Cell

Cas Protospacer Decoded

WANG ET AL., PAGE 840

The structure of the Cas1-Cas2-DNA complex reveals the dual-forked nature of the protospacer — the invading DNA sequence selected by Cas1 and Cas2 — and explains how its sequence and length are determined.

A Structural Core for Low Complexity

XIANG ET AL., PAGE 829

A chemical footprinting method reveals that polymers of low-complexity domains exhibit similar cross-β structure in hydrogels, liquid-like droplets, and in nuclei of mammalian cells, suggesting a common underlying structural basis.

Cascade Effects

REDDING ET AL., PAGE 854

Single-molecule analysis of the bacterial Cascade complex reveals how two distinct Cas pathways are engaged based on the presence or absence of PAM sequences in the DNA target, providing insights into how the CRISPR machinery adapts to mutations that allow foreign DNA to escape immunity.

Microtubules Mismanage Broken Chromosomes

LOTTERSBERGER ET AL., PAGE 880

Increased chromatin mobility at double-stranded break sites is mediated by nuclear envelope-associated proteins and microtubule dynamics, contributing to aberrant DNA repair.

Steroid Release on Steroids

YAMANAKA ET AL., PAGE 907

The *Drosophila* hormone ecdysone is released from cells through calcium-regulated vesicle trafficking, suggesting that steroid hormone release may be an active process rather than simple diffusion across the membrane.

Decorated Histone Veterans

XIE ET AL., PAGE 920

During stem cell division, a transient phosphate modification on histone H3 distinguishes pre-existing and newly synthesized histones and is required for the asymmetric segregation of sister chromatids—one enriched with new histones, the other with old.
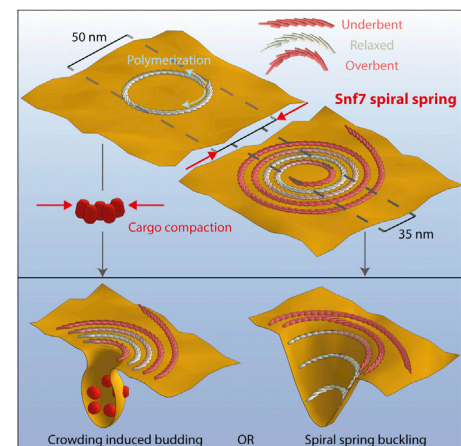
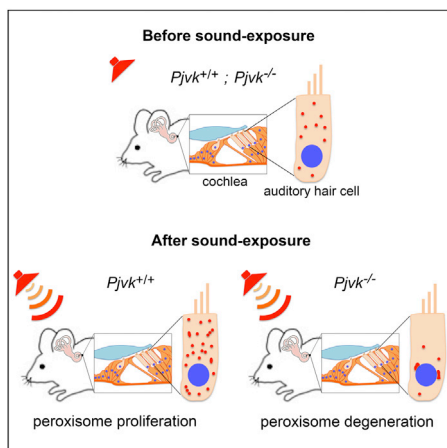## Polytene Bands = TAD

EAGEN ET AL., PAGE 934

Equating polytene bands with topologically associating domains (TADs) in interphase nuclei reveals two stable forms of folded chromatin within euchromatic regions of diploid cells that are distinct from more highly structured heterochromatin.

Springing Out of Shape

CHIARUTTINI ET AL., PAGE 866

A component of the ESCRT-III membrane fission machinery self-organizes into spiral springs that trigger membrane deformation when released.

Before sound-exposure

$Pjvk^{+/+}$; $Pjvk^{-/-}$

cochlea   auditory hair cell

After sound-exposure

$Pjvk^{+/+}$ $Pjvk^{-/-}$

peroxisome proliferation   peroxisome degeneration

Peroxisomes on Noise Patrol

DELMAGHANI ET AL., PAGE 894

Hypervulnerability to sound exposure in humans and mice with pejvakin mutations results from reduced adaptive proliferation of peroxisomes in response to the oxidative stress caused by loud sounds.

A TAD-bit to the Right

HU ET AL., PAGE 947

Chromosomal domains demarcated by CTCF and cohesin binding constrain RAG-mediated recombination and support a linear tracking model where breaks at defined motifs occur in a convergent orientation.

Endometriosis: The Beta Version

HAN ET AL., PAGE 960

Estrogen receptor β (ERβ) interacts with the cytoplasmic apoptotic machinery and the inflammasome complex to promote endometriosis—a significant cause of infertility.

Aire's Heir

TAKABA ET AL., PAGE 975

To promote immunological self-tolerance, Fezf2 directly regulates transcription of tissue-restricted antigen genes in the thymus, where it functions independent of the known self-tolerance regulator, Aire.

m^{6}A Coping Mechanism

MEYER ET AL., PAGE 999

N^6-methyladenosine (m^{6}A) modification is selectively increased in the 5′ UTR of mRNAs during cellular stress to promote translation initiation through a mechanism that does not require the 5′ cap or cap-binding proteins.

A Fine Look at Primary Prostate Cancer

SCHULTZ ET AL., PAGE 1011

Molecular analysis of 333 primary prostate carcinomas reveals substantial heterogeneity and major subtypes among patients, as well as potentially actionable lesions valuable for clinical management of the disease.
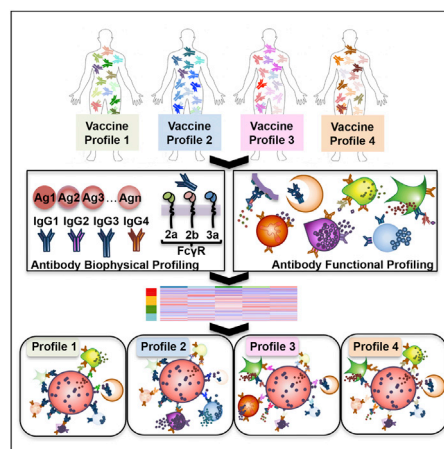
A Check on Cardiomyocyte Proliferation

ALKASS ET AL., PAGE 1026

Contrary to a recent report suggesting that a preadolescent burst of cardiomyocyte proliferation promotes heart growth, new evidence indicates that cardiomyocyte number expansion appears limited to the neonatal period, with cardiomyocyte hypertrophy likely accounting for the increase in the heart size.

Protective Fingerprints

CHUNG ET AL., PAGE 988

Systems serology reveals unique vaccine-induced "fingerprints," highlighting potential markers of protection against HIV and providing a powerful method for comparing candidate vaccines.

Vaccine Profile 1   Vaccine Profile 2   Vaccine Profile 3   Vaccine Profile 4

Ag1 Ag2 Ag3 ... Agn
IgG1 IgG2 IgG3 IgG4   2a 2b 3a
                          FcγR
Antibody Biophysical Profiling   Antibody Functional Profiling

Profile 1   Profile 2   Profile 3   Profile 4

Net Works for Malaria Control

In just a couple days, Youyou Tu, William C. Campbell and Satoshi Ōmura will meet His Majesty King Carl XVI Gustaf of Sweden to receive the Nobel Prize in Physiology or Medicine. Tu's discovery of artemisinin, a potent anti-malarial. and Campbell and Ōmura's work on ivermectin, a broad-spectrum antiparasitic drug, has helped to mitigate the burden of parasitic diseases, especially in the developing world. Despite all effort and advances, stopping the spread of preventable parasitic diseases remains an unmet goal. Bhatt et al. (Bhatt et al., 2015) now present the first data-driven comprehensive picture of how specific interventions for malaria control have impacted the spread of this disease in the past 15 years.

A demonstration of how to set up a mosquito bed net. Image courtesy of Andre Roussel, USAID.

In 1978, every 6 seconds one child would die from malaria. Today, one child still dies from malaria every minute. Malaria is a burden of approximately 600,000 deaths every year, concentrated largely in sub-Saharan Africa (Hemingway, 2015). First line treatment with the drug chloroquine and control of the exposure to the insect that transmits the disease with indoor spraying of the insecticide DDT failed, mostly due to acquisition of resistance by the Plasmodium—the malaria parasite, and the anopheline mosquitoes, the malaria vector.

The critical turning point in the combat against the disease was the launch of the Roll Back Malaria Initiative and the wider development agenda around the United Nations Millennium Development Goals, kicked off in the year 2000. The ambitious agenda: begin to reverse malaria incidence and to halt malaria spread by 2015. Since then, funding for malaria control has increased 20×, split between access to insecticide-treated bed nets, indoor residual spraying—actions to reduce exposure to the mosquitoes—prompt treatment of clinical malaria cases, and substitution of old drugs for highly efficacious artemisinin-based combination therapy.

Now that the benchmark year of 2015 has been reached, Bhatt et al. attempt to quantify the prevalence of *Plasmodium falciparum* infection and disease incidence across sub-Saharan Africa from 2000 to 2015, as well as to define the role major interventions have had in causing changes in malaria endemicity. By modeling disease transmission, the authors were able to generate counterfactual geospatial maps that provide estimation of what malaria parasite prevalence rates would look like today without each intervention. In total, they estimate that 663 million clinical cases of malaria were averted between 2000 and 2015. The distribution of bed nets alone was responsible for 68% percent of this improvement, followed by 22% resulting from artemisinin-based combination therapy, and 10% from indoor residual spraying. There are caveats to these numbers. For instance, they vary to some extent within different territories, and they are affected by how early and the scale that each intervention was deployed. Nonetheless, it may come as a surprise that such large fraction of the improvement can be attributed to interventions focused on mosquito-control.

Although the incidence of malaria has decreased to half of what it used to be 15 years ago, the new data suggest caution. Millions of people are still at risk of malaria disease and death in Africa, and rates of improvement slowed down about 5% per year in 2013. Continued distribution of bed nets, replacement of the old ones, surveillance for arousal of parasite and mosquito resistance—already documented in other areas in the world—will be essential steps to reduce the number of disease cases.

The World Health Organization and the Roll Back Malaria Partnership now have moved onto defining goals and priority actions for malaria control in the next 15 years period. Beyond providing an accurate picture of the effectiveness of malaria interventions in the recent past, the new data will be crucial to inform policy agencies with an optimal strategy for the future.

REFERENCES

Bhatt, S., Weiss, D.J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K.E., Moyes, C.L., Henry, A., Eckhoff, P.A., et al. (2015). Nature *526*, 207–211.

Hemingway, J. (2015). Nature *526*, 198–199.

João Monteiro

Cardiomyocyte Cell-Cycle Activity during Preadolescence

Earlier studies (Soonpaa et al., 1996) revealed a rapid drop-off of ventricular cardiomyocyte cell-cycle activity at birth in mice, followed by a burst of DNA synthesis during the first week of postnatal life which contributed to the formation of multi-nucleated cardiomyocytes by postnatal day 10 (PN10). It has recently been suggested that a second burst of cardiomyocyte cell-cycle activity occurs during preadolescence, between PN14 and PN18, resulting in a 40% increase in cardiomyocyte number (Naqvi et al., 2014). Since there was no overt change in mono- versus bi-nuclear cardiomyocyte content and no change in cardiomyocyte nuclear ploidy between PN14 and PN18, a 40% increase in ventricular cardiomyocyte number during preadolescence should result in newly synthesized DNA in 57% of the cardiomyocyte nuclei.

To characterize this putative burst of preadolescent cell-cycle activity, MHC-nLAC mice, expressing a nuclear-localized β-galactosidase reporter in cardiomyocytes and maintained in a DBA/2J background (Soonpaa et al., 1994), were implanted with BrdU-containing osmotic mini-pumps. Cumulative ventricular cardiomyocyte DNA synthesis was quantitated by co-localization of β-galactosidase and BrdU immune reactivity (Figure 1A) as described (Reuter et al., 2014). Only low levels of cardiomyocyte DNA synthesis were detected in mice carrying pumps from PN10 through PN19 (2.96% ± 0.55%) or from PN12 through PN19 (1.09% ± 0.33%; see also Table S1A). BrdU was detected in small intestine crypt cells by 24 hr postimplantation, and at the end of the labeling period (Figure 1B), confirming continuous infusion. Cumulative preadolescent cardiomyocyte DNA synthesis was also quantitated in C57Bl/6J inbred mice (the strain used by Naqvi and colleagues); S-phase cardiomyocytes were identified by nuclear BrdU immune reactivity in dispersed cell preparations (Figure 1C). Only low rates of cardiomyocyte DNA synthesis were detected (Table S1B). Mice receiving a single BrdU injection on PN14.5, PN15, or PN16 and analyzed on PN19 also had little labeling (Table S1C), arguing that BrdU cytotoxicity and/or the presence of the osmotic mini-pump per se were not confounding factors.
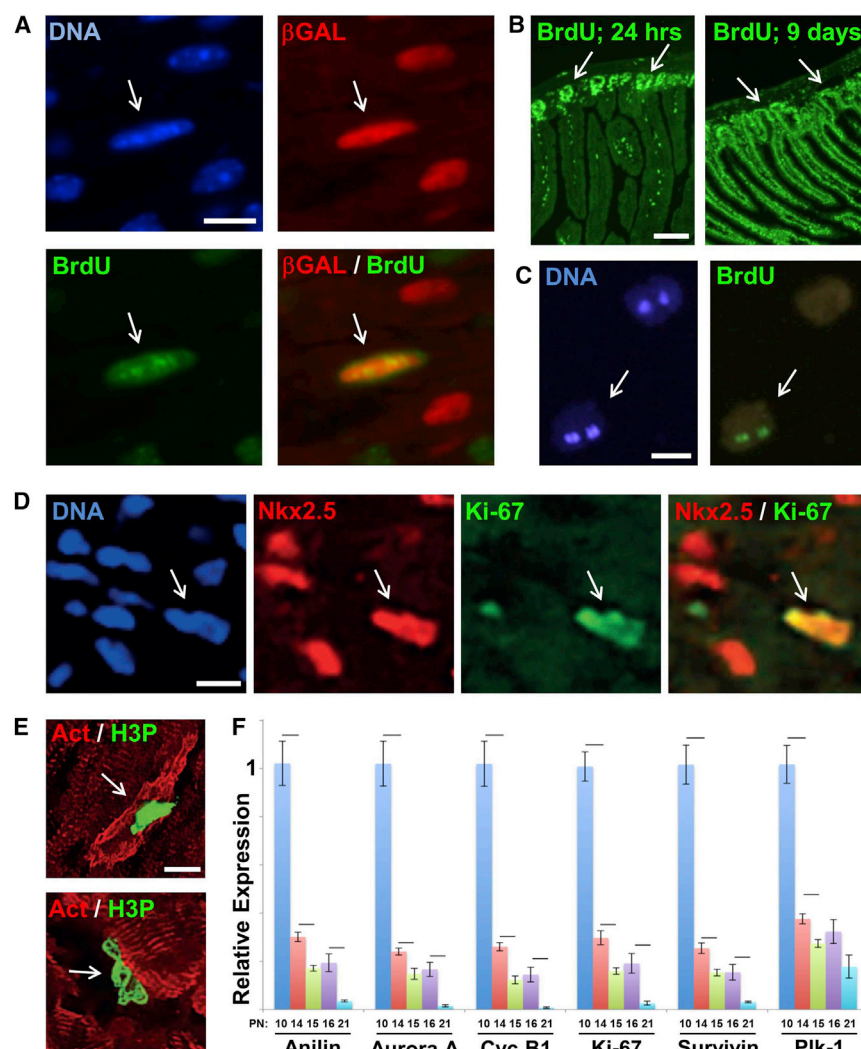

Figure 1. Characterization of Cardiomyocyte Cell-Cycle Parameters during Preadolescence

(A) Example of cardiomyocyte DNA synthesis (arrow) in the heart of an MHC-nLAC mouse carrying a BrdU mini-pump from PN10 through PN19. Panels show Hoechst staining of DNA (blue signal), beta-galactosidase immune reactivity (red signal) and BrdU immune reactivity (green signal). Scale bar, 10 microns.

(B) BrdU immune reactivity (green signal) in the small intestine of an MHC-nLAC mouse carrying a BrdU mini-pump for 24 hr (left) or 9 days (right). Arrows, villi crypts. Scale bar, 200 microns.

(C) Example of cardiomyocyte DNA synthesis (arrow) in dispersed cells from a C57Bl/6J mouse heart carrying a BrdU mini-pump from PN10 through PN19. Panels show Hoechst staining of DNA (blue signal) and BrdU immune reactivity (green signal). Scale bar, 20 microns.

(D) Example of an S-phase cardiomyocyte nucleus in a PN15 heart as evidenced by co-localization of Nkx2.5 (red signal) and Ki-67 (green signal) immune reactivity (arrow). Scale bar, 10 microns.

(E) Example of cardiomyocyte (upper panel) and non-cardiomyocyte (lower panel) H3P immune reactivity (green signal, arrow) in postnatal hearts. Cardiomyocytes were identified by α-actinin immune reactivity (red signal). Scale bar, 10 microns.

(F) Cardiac expression of mitosis-related genes on PN10, PN14, PN15, PN16, and PN21 in C57BL/6J mice (mean ± SEM) relative to their level at P10. mRNA levels were quantitated and normalized to 18S as described (Livak and Schmittgen, 2001). Significance was tested using unpaired, 2-tailed Student's t tests with Bonferroni correction for multiple testing; bars indicate p < 0.05 versus subsequent time point.

Cardiomyocyte cell-cycle activity was also quantitated via co-localization of Nkx2.5 and Ki-67 immune reactivity (Figure 1D); Ki-67 is expressed from G1 phase to anaphase and thus provides a very good estimate of the fraction of a given cell population with cell-cycle activity (Lopez et al., 1994). The cardiomyocyte nuclear Ki-67 labeling never exceeded 1% in C57Bl/6J ventricles between PN12 and PN16 (Table S1D), in agreement with the BrdU incorporation data. Phosphorylation of histone H3 on serine 10 (H3P), which labels cells from G2/M through early anaphase (Hendzel et al., 1997), was also used to monitor cardiomyocyte cell-cycle activity. Since cardiomyocyte mitosis is characterized by sarcomere disassembly (Engel et al., 2006), mitotic ventricular cardiomyocytes can be identified by the presence of H3P signal and sarcomere disassembly (Figure 1E). No mitotic cardiomyocytes were observed after PN12 (Table S1E).

Ventricular expression levels of a panel of mitosis-related genes (encoding anilin, aurora A, polo-like kinase 1, survivin, cyclin B1, and Ki-67) were measured by qPCR in C57Bl/6J mice, and the relative levels of expression were normalized to that of 18S rRNA as described (Liu et al., 2015). No changes in transcript levels supporting the presence of a proliferative burst between PN14 and PN16 were detected (Figure 1F). Collectively, the low rates of cardiomyocyte DNA synthesis (analyzed by M.H.S. and L.J.F.), the low levels of cardiomyocyte Ki-67 and H3P immune reactivity (analyzed by D.C.Z. and F.B.E.) and the absence of transient increases in mitotic transcripts (analyzed by C.P and A.R) during preadolescence are consistent with the pattern of gradual postnatal ventricular cardiomyocyte cell-cycle withdrawal reported earlier using single injections of tritiated thymidine to monitor cardiomyocyte cell-cycle activity (Soonpaa et al., 1996).

Potentially trivial factors such as the method for defining postnatal age, strain differences, BrdU cytotoxicity, cell identification in histologic samples, and the sequences of the PCR primers used, were controlled for and thus cannot explain the differences in the results presented here and those of Naqvi and colleagues. Whether subtle differences in animal husbandry can explain the differences in the measured parameters between the two studies remains unclear and worthy of further investigation. Indeed, the current study does not rule out the possibility that a preadolescent burst in cardiomyocyte cell-cycle activity can exist, assuming that as of yet undefined optimal conditions of litter size, nutrients, etc., are met. However, the data presented here clearly indicate that preadolescent development did occur in the absence of a burst of cardiomyocyte proliferation over multiple litters in three independent breeding colonies. Consequently, a burst of preadolescent cell-cycle activity is not required for normal cardiac development.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one supplemental table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.037.

Mark H. Soonpaa,[1]
David C. Zebrowski,[2] Colin Platt,[3]
Anthony Rosenzweig,[3] Felix B. Engel,[2]
and Loren J. Field[1,*]
[1]The Krannert Institute of Cardiology, and the Riley Heart Research Center, Wells Center for Pediatric Research, and Indiana University School of Medicine, 1044 West Walnut Street, Indianapolis, IN 46202, USA
[2]Experimental Renal and Cardiovascular Research, Department of Nephropathology, Institute of Pathology, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Schwabachanlage 12, 91054, Erlangen, Germany
[3]The Cardiovascular Division of the Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02215, USA
*Correspondence: ljfield@iupui.edu
http://dx.doi.org/10.1016/j.cell.2015.10.037

REFERENCES

Engel, F.B., Schebesta, M., and Keating, M.T. (2006). J. Mol. Cell. Cardiol. 41, 601–612.

Hendzel, M.J., Wei, Y., Mancini, M.A., Van Hooser, A., Ranalli, T., Brinkley, B.R., Bazett-Jones, D.P., and Allis, C.D. (1997). Chromosoma 106, 348–360.

Liu, X., Xiao, J., Zhu, H., Wei, X., Platt, C., Damilano, F., Xiao, C., Bezzerides, V., Boström, P., Che, L., et al. (2015). Cell Metab. 21, 584–595.

Livak, K.J., and Schmittgen, T.D. (2001). Methods 25, 402–408.

Lopez, F., Belloc, F., Lacombe, F., Dumain, P., Reiffers, J., Bernard, P., and Boisseau, M.R. (1994). Exp. Cell Res. 210, 145–153.

Naqvi, N., Li, M., Calvert, J.W., Tejada, T., Lambert, J.P., Wu, J., Kesteven, S.H., Holman, S.R., Matsuda, T., Lovelock, J.D., et al. (2014). Cell 157, 795–807.

Reuter, S., Soonpaa, M.H., Firulli, A.B., Chang, A.N., and Field, L.J. (2014). PLoS ONE 9, e115871.

Soonpaa, M.H., Koh, G.Y., Klug, M.G., and Field, L.J. (1994). Science 264, 98–101.

Soonpaa, M.H., Kim, K.K., Pajak, L., Franklin, M., and Field, L.J. (1996). Am. J. Physiol. 271, H2183–H2189.

Cardiomyocytes Replicate and their Numbers Increase in Young Hearts

In mice, the heart nearly quadruples in size from early preadolescence (postnatal day 10 [P10]) to puberty (∼P35) (Naqvi et al., 2014). Because it is widely believed that mammalian cardiomyocytes (CMs) are incapable of replication after birth, it has been assumed that early postnatal heart growth is driven solely by CM hypertrophy. Our findings question this view and provide insights into how thyroid hormone may regulate an increase in CM number during preadolescence. Alkass et al. (2015) and Soonpaa et al. (2015) present evidence that conflicts with our conclusions; we believe that differences in experimental design and methodology could account for the discrepancy.

Alkass et al. (2015), using design-based stereology, concluded that the murine CM population does not increase during preadolescence. Their measurements of CM numbers (Figure S1C in Alkass et al.), when pooled into groups encompassing P5–P9 and P11–P100, show an ∼20% increase between the two developmental periods (p < 0.001), indicating a CM population expansion at some time after P9 (or early preadolescence). But the authors, based on comparisons of CM numbers on single days, conclude that there was no increase during preadolescence (Figure 1E in Alkass et al.). This conclusion may stem from the large variance in their data between P11 and P100, together with a low number of replicates. Given the reported SD of means (for CM numbers) between P11 and P100, and with three replicates, their study is underpowered to reliably refute even a 40% change in CM numbers.

In agreement with our findings, analysis of raw data

from a study (Puente et al., 2014) that used cell disaggregation/counting protocols to determine the murine CM population (Figures 4G and 7I of Puente et al. [2014], data kindly supplied by Hesham Sadek), showed an increase of 33% between P7 and P21 (p = 0.0396), and of 29% between P7 and P14 (p = 0.029) (Student's t test, two-tailed).

Aurora B, cTNT, DAPI

Figure 1. Verifying CM Mitoses in P15 Mouse Cardiac Left Ventricle by an Independent Laboratory from Blinded Histological Heart Samples
(A and B) Immunohistochemical identification of mitotic CMs (red staining, Aurora B-positive cells) in transverse cut tissue sections showing localization in the left ventricle (LV) of P15 mice (B), with no Aurora B-positive cells evident in the LV of the P10 heart (A). CMs are labeled with cardiac troponin T (cTNT) (green) and nuclei are labeled with DAPI (blue). Arrows show some Aurora B-positive CMs. Experimental procedures used for immunohistochemistry were essentially as described (Naqvi et al., 2014). The Aurora B antibody (Abcam, ab 2254), supplied as a 1-mg/ml solution, was used at a dilution of 1:50. Sections were prepared and antibodies provided by the Graham laboratory. Immunohistochemical staining was performed in a blinded manner by the Harvey laboratory.

Our counting method relies on similar tissue digestion/disaggregation efficiencies between hearts of different ages. To validate our findings, we calculated CM numbers in the heart by dividing total ventricular volume (occupied by CMs) by the average CM volume (Naqvi et al., 2014), an approach independent of digestion efficiency. This also revealed an increase in CM numbers during preadolescence, supporting the numbers estimated by direct cell counting. Alkass et al. offer the unsupported assertion that apparent increases in the CM population are caused by variations in digestion efficiency between hearts of different ages, but this is refuted by our finding of markedly reduced CM numbers in mice of the same age when T3 biosynthesis is inhibited (Naqvi et al., 2014).

Design-based stereology, used by Alkass et al., relies on accurate delineation of CM boundaries over several tissue planes; it is ideal for nearly spherical neonatal CMs, but technically challenging when CMs become elongated and branched during preadolescence. Moreover, Alkass et al. randomly sampled only 1%–2% of the heart, which they assume to be homogeneous. We found that CMs undergoing mitosis and cytokinesis are present mainly in the subendocardium (Naqvi et al., 2014), indicating cellular heterogeneity in thyroid hormone response. Hence, random assessment of a few sites, as in Alkaas et al., is likely to increase estimation errors. Together with analysis of only a few hearts at each age, this could produce errors that are reflected in a very high variance in their CM numbers, and thereby obscure biologically meaningful increases in CM numbers during preadolescence.

Alkass et al. and Soonpaa et al. were unable to detect mitoses in preadolescent CMs. We thus sent unprocessed murine heart sections to Richard Harvey's laboratory (Victor
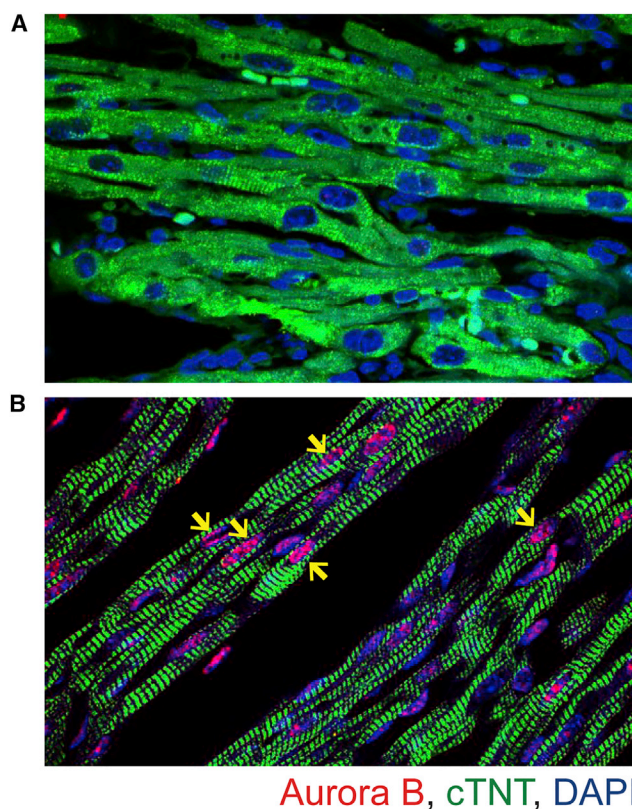
Chang Cardiac Research Institute). Blinded experiments indicated nuclear Aurora B labeling (which marks mitotic cells) in hearts at P2, P14 (harvested 9 p.m.), and P15, but not P10 or P18 (Figure 1A and 1B and data not shown), consistent with Naqvi et al. (2014). Soonpaa et al. declined our offer of unprocessed tissue sections from Naqvi et al. (2014) for independent validation. Alkass et al. dispute that Aurora B+ nuclei (Figure 3B in Naqvi et al. [2014]) are in CMs. However, we showed unambiguous sarcomeric α-myosin heavy chain labeling in isolated Aurora B+ P15 CMs (Figure 3G in Naqvi et al. [2014]), which is also evident in a very high-resolution micrograph of a P15 heart section (Figure S1). These data confirm mitosis in these CMs.

Both Alkass et al. and Soonpaa et al. invoke a discrepancy between our estimate of the number of new CMs born at ∼P15 (direct counting) and estimates of the percentage of CMs that have undergone a new S-phase (BrdU labeling). Our numbers are similar to those independently observed by Murray et al. (2015) using a protocol like ours and are consistent with extensive other data we have published (Naqvi et al., 2014). We suggest that S-phase CMs are underestimated by the methods used by Alkass et al. and Soonpaa et al. We are not surprised that a BrdU pulse on P14 labels only ∼3%–5% of CMs (Naqvi et al., 2014; Murray et al., 2015); in rodents the time of tracer clearance from blood ranges from 0.5 to 1 hr, short compared with the length of S-phase (Duque and Rakic, 2011), and S-phase in CMs may not be synchronized. While continuous BrdU infusion avoids this limitation, the anti-proliferative effects of BrdU/EdU could depress CM division, and toxicity may increase death of labeled cells. Thus the BrdU/EdU labeling index, after pulse or continuous administration, is not an infallible indicator of the extent of cell

replication (Duque and Rakic, 2011). CM numbers after birth vary strikingly with litter size (Bai et al., 1990), and variations in animal husbandry, litter size, or gender could affect both the timing and time-to-peak of environmentally programmed hormonal influences on CM numbers; thus we have analyzed only male C57BL/6J litters of 6–7 pups with confirmed birth dates.

In summary, the findings of Alkass et al. and Soonpaa et al. do not refute our multiple lines of evidence indicating an increase in the CM population during preadolescence (Naqvi et al., 2014). In support of our findings, Wulfsohn et al. (2004), using stereology, found a doubling of CM numbers in rats from P25–P125, and Mollova et al. (2013), also using stereology, found a 3.4-fold increase in human CM numbers from 1–20 years. These data, and our evidence for markers of CM mitoses, mitotic figures, acute decreases in cell volume consistent with CM replication, and BrdU labeling, provide compelling evidence for CM proliferation during preadolescence (Naqvi et al., 2014). Finally, Mollova et al. (2013) and Bergmann et al. (2009) found that most human CMs are formed during the first 10 years of life—that is, in preadolescence—which is the key finding of Naqvi et al. (2014).

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figures and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.038.

Nawazish Naqvi,[1] Reena Singh,[2] Siiri E. Iismaa,[3] Ming Li,[3] John W. Calvert,[4] David I.K. Martin,[5] Richard P. Harvey,[2] Robert M. Graham,[3,*] and Ahsan Husain[1,*]
[1]Department of Medicine, Emory University, Atlanta, Georgia 30322, USA
[2]Development and Stem Cell Biology Division
[3]Molecular Cardiology and Biophysics Division
Victor Chang Cardiac Research Institute, Sydney, NSW 2010, Australia
[4]Department of Surgery, Emory University, Atlanta, Georgia 30322, USA
[5]Children's Hospital Oakland Research Institute, Oakland, California 94609, USA
*Correspondence: ahusai2@emory.edu (A.H.), b.graham@victorchang.edu (R.M.G.)
http://dx.doi.org/10.1016/j.cell.2015.10.038

REFERENCES

Alkass, K., Panula, J., Westman, M., Wu, T.-D., Guerquin-Kern, J.-L., and Bergmann, O. (2015). No Evidence for Cardiomyocyte Number Expansion in Preadolescent Mice. Cell 163, this issue, 1026–1036.

Bai, S.L., Campbell, S.E., Moore, J.A., Morales, M.C., and Gerdes, A.M. (1990). Anat. Rec. 226, 207–212.

Bergmann, O., Bhardwaj, R.D., Bernard, S., Zdunek, S., Barnabé-Heider, F., Walsh, S., Zupicich, J., Alkass, K., Buchholz, B.A., Druid, H., et al. (2009). Science 324, 98–102.

Duque, A., and Rakic, P. (2011). J. Neurosci. 31, 15205–15217.

Mollova, M., Bersell, K., Walsh, S., Savla, J., Das, L.T., Park, S.Y., Silberstein, L.E., Dos Remedios, C.G., Graham, D., Colan, S., and Kühn, B. (2013). Proc. Natl. Acad. Sci. USA 110, 1446–1451.

Murray, T.V., Smyrnias, I., Schnelle, M., Mistry, R.K., Zhang, M., Beretta, M., Martin, D., Anilkumar, N., de Silva, S.M., Shah, A.M., and Brewer, A.C. (2015). J. Mol. Cell. Cardiol. 79, 54–68.

Naqvi, N., Li, M., Calvert, J.W., Tejada, T., Lambert, J.P., Wu, J., Kesteven, S.H., Holman, S.R., Matsuda, T., Lovelock, J.D., et al. (2014). Cell 157, 795–807.

Puente, B.N., Kimura, W., Muralidhar, S.A., Moon, J., Amatruda, J.F., Phelps, K.L., Grinsfelder, D., Rothermel, B.A., Chen, R., Garcia, J.A., et al. (2014). Cell 157, 565–579.

Soonpaa, M.H., Zebrowski, D.C., Platt, C., Rosenzweig, A., Engel, F.B., and Field, L.J. (2015). Cardiomyocyte Cell-Cycle Activity during Preadolescence. Cell 163, this issue, 781–782.

Wulfsohn, D., Nyengaard, J.R., and Tang, Y. (2004). Anat. Rec. A Discov. Mol. Cell. Evol. Biol. 277, 236–247.

A Gut Feeling about You

Follow Your Gut: The Enormous Impact of Tiny Microbes
Authors: Rob Knight with Brendan Buhler
TED Books and Simon & Schuster: New York, NY, USA (2015).
114 pp. $16.99

The "gut microbiota" is rapidly becoming a common term outside of the halls of science: it has been headlined in the *New York Times*, is the subject of several non-fiction books, and is regularly promoted on TV (probiotics, anyone…?). Many non-scientists may not remember or know the technical term microbiota, but mention the gut flora to them, and they'll probably know what you're referring to.

Every month, new research comes out describing yet more findings about the gut microbiota or, more broadly, the human microbiota—your collection of microbes across all body locations. With each new study, more intriguing questions unfold as scientists try to understand the mechanisms involved. In this short TED book, Rob Knight from the University of Boulder, Colorado, one of the main researchers involved in the Human Microbiome Project, and Brendan Buhler, an award-winning science writer, provide a knowledgeable, up-to-date summary of how your gut microbiota affects many aspects of your everyday life. "Allergies, asthma, obesity, acne: these are just a few of the conditions that may be caused—and someday cured—by the microscopic life inside us. The key is to understand how this groundbreaking science influences your health, mood, and more," state the authors. And they are right.

In seven short chapters, we learn a myriad of facts about the gut microbiome and ourselves. Starting with the first chapter, we discover how microbial we are; in fact, humans are mostly composed of microbes. The book then covers how we acquire our microbiome, its emerging role in sickness and in health, how it interacts with our brain through the fascinating gut-microbiome axis, and how we might be able to tailor it to our specific needs. The book also touches upon the important topic of antibiotics and their effects on the gut microbiota before closing with a chapter looking to the future, as envisioned by the authors.

Overall, the studies referenced are solid, and the often-personal anecdotes throughout the book are generally spot-on and funny. We learn of the sobering arrival of Rob Knight's daughter into the world, how our pets resemble us more than we think, and how bacteria could help us stop worrying about our waistlines. The reading is easy and enjoyable, with simple cartoons illustrating the different topics at hand. Two sidebars broaden the scope of the book, providing us with a "Brief History of Bugs," which

covers the days from Antonie van Leeuwenhoek's first observations of bacteria up to when Robert Koch linked them to disease with his famous postulates, and "The Science (and Art) of Microbiome Mapping," which provides a window into how researchers go about decoding the genomic content from complex collections of microbes. The book also contains an addendum on "The American Gut," an ongoing open-source scientific project led by Dr. Knight, allowing each of us to discover which inhabitants make up our own individual microbiota. Thousands of donations have already been made and sequences are publicly available, allowing researchers to map the incredible diversity of the gut microbiome and identify trends that can be further investigated in the lab.

For those already involved in the field of human microbiome research, this book will not provide anything new. Of particular interest, though, is how Rob Knight compares your gut microbiota to a garden. With this analogy, he makes the very strong point that microbial ecology hypotheses and testing will be crucial if we are to fully understand the extent to which our microbiota influences our health.

For those who wish to learn more about this exciting field, this will be an easy first step. And for those seeking health advice involving their gut microbiota, Dr. Knight provides wise advice: check the sources. Before making any radical change in your lifestyle or believing the overwhelming claims about a miracle-microbial solution, you should ask yourself, "Who says so, and how does he or she know?" This book provides helpful hints for an evidence-based approach to evaluating the various health claims out there.

In short, this enjoyable read summarizes a rather complex field of science with some everyday scenarios and delivers an optimistic, yet not unrealizable, vision of the future of medicine. Clearly, we have much to learn about ourselves as individual microbial ecosystems.
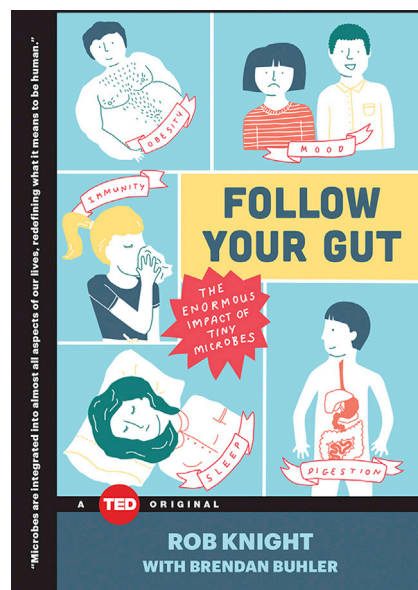
Corinne F. Maurice[1,*]
[1]McGill University, Department of Microbiology and Immunology, Microbiome and Disease Tolerance Centre, Montreal H3A 2B4, Canada
*Correspondence: corinne.maurice@mcgill.ca

Adopting Rare Diseases

Orphan: The Quest to Save Children with Rare Genetic Disorders
Author: Philip R. Reilly
Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA (2015).
369 pp. $29.00

In the United States, rare diseases are defined as conditions affecting fewer than 200,000 individuals at any given time. Worldwide, an estimated 300 million people are affected by rare diseases, also known as "orphan diseases" because with few financial incentives for developing new treatments, the pharmaceutical industry did not "adopt" them. In his book entitled *Orphan,* Phillip Reilly paints a picture of several rare diseases and the quest to find treatments for them.

Recent years have witnessed a sea change in attention to rare diseases due to the convergence of regulatory, scientific, and societal forces. In 1983, the US Congress enacted the Orphan Drug Act (ODA), providing a number of incentives for companies to work on rare diseases. In addition, the blockbuster drug model focusing on common, highly profitable conditions such as hypertension or hypercholesterolemia proved to be unsustainable for the pharmaceutical industry. Many companies have instead turned to rare diseases, especially those that are genetic in origin and have a well-characterized biological mechanism. Finally, the technologies necessary to intervene in genetic disorders are starting to mature, enabling a number of therapeutic options. It is amidst this synergistic set of circumstances that Reilly's book arrives, describing the advances for some of these disorders, the heroes (parents, scientists, clinicians, companies) behind these advances, and the work left to be done.

Reilly brings a broad and unique perspective on these disorders to his book. With both MD and JD degrees to his credit and board certification in Internal Medicine and Clinical Genetics, he joined the Eunice Kennedy Shriver Center for Mental Retardation in the mid-1980s. There, he provided primary care for 800 adults in a state-run institution for people with severe neurological conditions. He then joined a venture capital firm in 2008 and has been focusing on starting companies to treat rare genetic disorders. This wide range of experiences allows him to move with relative ease between describing the impact of these disorders on individuals and their families to discussing the societal and economic impacts of decisions surrounding these conditions. His work as a physician, geneticist, and venture capitalist, together with his evident interest in history of medicine, contribute to his narrative of the thrilling advances in the area of rare diseases and how they have touched people's lives.

Reilly observes, "To build a company with the goal of developing a novel, transformative therapy for an orphan genetic disorder one must understand the disorder and its impact on families at a level that can never be attained simply by reading medical journals." Accordingly, *Orphan* is full of stories about individuals Reilly has gotten to know under various circumstances. The most vivid vignettes are those of families affected with rare diseases such as phenylketonuria (PKU), a disease that has come under control with dietary treatments; dystrophic epidermolysis bullosa, a life-threatening skin condition; and X-linked adrenoleukodystrophy, a neurological disorder presenting with cognitive decline in elementary school years. Since ~50% of the people affected by rare diseases are children, many of the vignettes highlight their parents. Parents have made enormous contributions to the advances in rare disease treatment by being advocates, lobbyists, and entrepreneurs—but above all, courageous caregivers willing to enroll their children in phase I or II clinical trials.

Other heroes in the book are physicians and scientists who focus on a rare disorder and make it their career to seek a treatment. One distinguished example among many is Roscoe Brady, a physician scientist and outstanding biochemist at the National Institute of Neurological Disorders and Stroke (NINDS). Brady focused on Gaucher disease, a disorder in which the lipid glucocerebroside accumulates in the body, leading to liver and spleen enlargement. Brady and colleagues discovered that the synthesis of this lipid was normal, but its clearance by a key lysosomal enzyme was impaired in people with Gaucher disease. In 1966, 20 years before the gene encoding this enzyme was cloned, Brady proposed treating Gaucher patients with enzyme replacement, but it was not until 1991 that the success of such an approach was demonstrated in a clinical trial using enzyme extracted from placentas. Since then, a recombinant protein replacement approach has been used effectively to treat a number of diseases of lysosomal enzyme deficiency.

Over the last two decades, similar approaches have been approved for a number of rare diseases, but patients must take these medications regularly for the rest of their lives. Newer approaches using gene therapy are trying to change that, and Reilly discusses the ups and downs of this research in detail. Gene therapy is experiencing a renaissance fueled by advances in viral vector technology and is being tested in many ongoing clinical trials, including for rare diseases such as spinal muscular atrophy, the
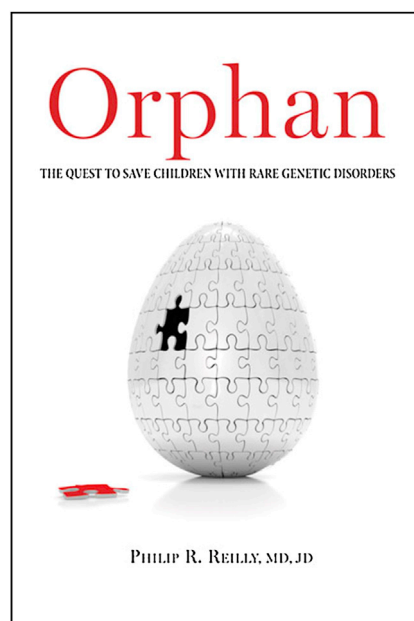
leading genetic cause of infant mortality. To date, only one gene therapy product (Glybera, a drug to treat a rare lipid disorder) is approved in Europe; none are approved in the US. Nonetheless, there is great promise for viral-mediated gene delivery to turn diseases with high mortality into chronic conditions.

In addition to viral gene delivery, other technologies are likely to expand therapeutic possibilities, reduce side effects, and improve efficacy in rare diseases. Reilly summarizes these advancements, such as exon skipping, RNA interference, induced pluripotent stem cells, and gene editing. According to a 2013 American Biopharmaceutical Research Companies report, more than 450 drugs are in development for rare diseases. While this is undoubtedly significant progress, there are still scientific, regulatory, and economic hurdles to drug development for rare diseases. One major scientific obstacle remains the blood-brain barrier, which is impenetrable to many small molecules and almost all biologicals. Given that many rare genetic diseases affect the nervous system, technologies that can traverse the blood-brain barrier are a high priority. In addition, for many diseases, we still do not know how early treatment must be initiated to be effective, so defining the temporal windows critical for treatment is a major challenge. Moreover, clinical trials may be difficult to design for rare diseases due to small patient populations, variability of expression, and lack of validated endpoints, especially for those affecting the central nervous system. These hurdles can be further complicated by diverse mutations in the same disease gene that sometimes require different treatments, as exemplified by recent treatments developed for cystic fibrosis patients carrying only certain mutations. While we have reason to be enthusiastic about the promise of mechanism-based treatments, we need to be realistic about these challenges so that we can overcome them.

"We are all orphans" is the provocative title of Reilly's final chapter. Copy-number variants and whole-exome analysis studies have revealed that each one of us carries rare changes in our genome that could affect our health. For many of them, we do not yet know whether there is any clinical significance, but some common diseases will likely be caused by a combination of genetic variants, as we are discovering for autism spectrum disorder. Highly penetrant genetic variations may contribute to some of the more than 7,000 diseases that are considered rare diseases, affecting an estimated 20–30 million Americans. The Precision Medicine Initiative put forth by President Obama may therefore have the largest initial impact on rare diseases.

Reilly acknowledges the societal, ethical, and financial considerations in this era of huge progress in rare disease research and treatment. Increased detection during gestation and the potential necessity of treatment at fetal stages raise major ethical issues. As genetic diagnosis technology improves, there will be increased demand for Masters-level trained genetic counselors, who are already scarce in our healthcare system. As effective therapies become available, there will be fewer reasons to deny genetic diagnostic tests to patients. Moreover, the high cost of certain approved treatments has already attracted attention. Some of the newer therapies will turn chronic costs into large one-time treatment costs. How will such costs be reimbursed? What will the impact be in the developing world? Currently, it is estimated that a quarter of the beds in pediatric hospitals are occupied by patients with genetic diseases. As a child neurologist investigating some of these rare diseases, I suspect that number is underestimated, not accounting for many undiagnosed genetic disorders.

"This book is about the struggle to save the lives of children who, because of a roll of the genetic dice, are born with any one of the more than several thousand rare genetic disorders." In *Orphan,* Reilly skillfully describes the promise and challenges of treating rare diseases and encourages public discourse about where we can be in the next few decades. The book will likely engage not only those working on or affected by rare diseases, but also students and biomedical scientists broadly. Different readers may be drawn to different aspects of the book: resilience of the individuals and families living with these conditions, progress in basic science and technology, or the due diligence process for a startup company. A combination of our resources, participation from families, hard work from physicians and scientists, and the support of industry and funders will be needed to translate genetic discoveries into treatments that will improve the lives of those touched by rare diseases.

Mustafa Sahin[1,*]
[1]Translational Neuroscience Center, FM Kirby Neurobiology Center, Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA
*Correspondence: mustafa.sahin@childrens.harvard.edu

Radical Sabbaticals

Whirlwind World Tour

Hans Clevers
Hubrecht Institute

Growing up, I believed that academics were shielded from the clutter of everyday life and had the peace to think, discuss, perform the occasional experiment and think some more. After starting a research lab, however, I discovered there was not much peace in academia. The day was just not long enough to fit everything in, and though a sabbatical leave always lurked at the horizon, there never seemed to be a good moment to take a break.

After three hectic years at the helm of the Royal Netherlands Academy and a particularly productive period for the lab (in my virtual absence, as the postdocs like to point out), a sabbatical could no longer be postponed. With two sons at university, I did not need to stay in one place for an entire year. My years at the Academy made me aware of how profoundly individual scientists and their science are affected by their environment, so my sabbatical became a world tour: six weeks each in six countries, interspersed by four week stints back in Utrecht, with great university cities serving as a home base from which to travel to teach and talk science. From Melbourne, I visited Sidney, Adelaide, and Brisbane, and from the Weizmann, I went to Tel Aviv, Haifa, and Jerusalem. New York is next, then San Diego and Paris, and finally Hong Kong. The experience will be a unique opportunity to see what drives scientists today, where science is going, and what the local and global challenges are.

My "Big Year"

Stuart Firestein
Columbia University

Four years ago, an exceptional undergrad in my lab asked me for a letter of reference for a Masters program in the History and Philosophy of Science (HPS) at Cambridge University. When I read about the program, my first thought was, "Hell, *I'd* like to do this." At the time, I was just finishing a book about how our ignorance fuels science. Three years later, having completed my sentence…er, term as department Chair, I contacted HPS and asked to join their Masters program during my sabbatical. They felt I might be "a bit too senior" for that, but suggested instead that I come as a Visiting Scholar, an offer I couldn't refuse.

Many scientists may consider philosophers and historians of science to be like birdwatchers, impacting scientists about as much as ornithologists affect birds. That may be true for benchwork, but a key part of our job is shaping the process and trajectory of science. Our perspective on these matters so crucial to the success of science is generally limited to our own meager individual experience. At HPS, I was exposed to the rigorous scholarship and broad view of science historians. I learned, among other things, how turn-of-the-century physics imbedded itself into the philosophical and social thinking of the culture. Might biologists today be missing a similar opportunity to engage citizens? At an age when one would think my most affecting experiences were behind me, I had the most important year of my life at HPS.

Embracing the Whiteboard

Leonie Ringrose
Humboldt University of Berlin

Last winter, I found myself in the unusual position of having some time on my hands. My Junior PI position at IMBA Vienna had come to an end, and I was still negotiating my senior position at Humboldt University in Berlin. Ever since a brief fling with kinetic mathematical modeling during my Ph.D. in the 1990s, I've been wanting to bring the crisp formalism of mathematical modeling to the perplexing questions of epigenetics but never seemed to find the time to pursue this. So when I met Martin Howard, a physicist at the John Innes Centre in Norwich who uses stochastic modeling to study epigenetic memory in plants, I was inspired by his approach and managed to talk him into hosting me for a few months to see if we could adapt his models to questions in *Drosophila* epigenetics.

Once there, I found it wonderfully refreshing to have the sole task of working on my project. I overcame my initial fear of giving group meetings with no PowerPoint slides, just a whiteboard. The project was a success: the combination of Martin's expertise and interest in the topic, the environment of theoretical thinkers, combined with my background in the details of the biology, turned out to be very fruitful. I thoroughly benefitted not just from taking a career break but also from leaving my comfort zone and taking a leap into the unknown. And I will certainly be making more use of modeling, and of the whiteboard, in the future.

Rescuing Abandoned Drugs

Rene Bernards
Netherlands Cancer Institute

After studying cancer genetics for 23 years, I wanted to get a closer look at how pharma approaches cancer drug discovery. To do this, I spent a 9 month sabbatical at the Discovery Oncology Department at Genentech in San Francisco. My first surprise was that the research atmosphere there is rather academic, with small research groups led by accomplished scientists. They were very generous and quickly made me feel at home. I was especially interested to learn how decisions are made about which molecules are taken forward in clinical development. This led to the second surprise, namely that, in pharma, it is practically a mantra that drugs must have demonstrated single agent activity before drug combinations are considered. Due to the extensive redundancy and feedback loops in the major signaling pathways in cancer, many good drugs may be abandoned early on for lack of such single agent activity. As one example, had the highly successful melanoma drug vemurafenib (a BRAF inhibitor) first been tested in *BRAF* mutant colon cancer, the drug would have failed miserably, as it only shows activity in this context in combination with EGFR inhibition.

So the take-home lesson from my sabbatical is that some perfectly good drugs may have been needlessly abandoned and might be resurrected from their graves with synthetic lethal screens to find effective drug combinations. It was time well spent, and I came away with lots of new ideas to work on!

TB for Two

K. Heran Darwin and Russell E. Vance
New York University Medical Center and University of California, Berkeley (HHMI)

We are two biologists, from opposite sides of the country and distinct training backgrounds, but united in our interest in infectious diseases. Heran's lab focuses on the genetics of *Mycobacterium tuberculosis* pathogenesis, whereas Russell's lab focuses on innate immune responses to infection. *M. tuberculosis* infects more than 2 billion people worldwide, causing over a million deaths annually. There is no effective vaccine, and the bacterium is highly contagious, requiring special BSL3 procedures to be safely studied in the lab. On top of this, tuberculosis (TB) is a complex disease where both the pathogen and the host are complicit in causing mortality.

Because TB and immunology are both fields that are difficult for outsiders to enter, we realized that we would both benefit from a "mutual" sabbatical. After discussions starting back in 2012, and the intervening chaos of Superstorm Sandy that disrupted Heran's lab at NYU, we were finally able to arrange for Heran to visit Berkeley for a sabbatical to coincide with Russell's "stay-batical" this fall/spring. To help facilitate interaction, we are sharing a single office at Berkeley. Our immediate goals are for Russell to get in the lab and learn how to work with TB, and for Heran to think more about TB from the perspective of the host. Our longer-term ambitions are to try to bring together our knowledge of microbiology and immunology to make progress on this difficult and globally challenging disease.

Peroxisomes Get Loud:
A Redox Antidote to Hearing Loss

Pablo Mardones[1,2] and Claudio Hetz[1,2,3,4,*]
[1]Biomedical Neuroscience Institute, Faculty of Medicine, University of Chile, Santiago, Chile
[2]Program of Cellular and Molecular Biology, Center for Molecular Studies of the Cell, Institute of Biomedical Sciences, Faculty of Medicine, University of Chile, Santiago, Chile
[3]FONDAP Center for Geroscience, Brain Health and Metabolism, University of Chile, Santiago, Chile
[4]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston MA 02115, USA
*Correspondence: chetz@hsph.harvard.edu
http://dx.doi.org/10.1016/j.cell.2015.10.060

Pejvakin (PJVK), a protein originally identified in Persian families with sensorineural hearing loss, regulates peroxisomal dynamics and the antioxidant defense triggered by noise exposure in hair cells and auditory neurons of the inner ear. These findings bring peroxisomes to the forefront of noise-induced hearing loss research.

The ability to respond to auditory cues is a highly prized evolutionary innovation in vertebrates, providing significant advantages for communication and the perception of environmental stimuli. The mammalian inner ear is a particularly intricate structure, engineered through natural selection for the detection of a wide range of sound frequencies and energies. Transduction of air pressure waves into meaningful sounds relies on a neural interface provided by highly specialized hair cells in the cochlea that transmit mechanical information to primary auditory neurons with remarkable precision for subsequent decoding in the auditory cortex (Figure 1A). As with all nervous tissue, the ability of hair cells to withstand damage and self-repair is severely limited. In particular, their exposure to high-energy sounds (>100–120 dB) results in mechanical trauma leading to irreparable structural damage and permanent hearing loss. Prolonged exposure to lower thresholds of high-energy sound (>85 dB) is a major cause of oxidative stress in hair cells, which may also lead to cell death and hearing loss (Wong and Ryan, 2015). Indeed, noise-induced hearing loss (NIHL) is a main cause of auditory disability and one of the most prevalent occupational hazards (Nelson et al., 2005). In this issue of *Cell*, Delmaghani et al. (2015) show that Pejvakin (PJVK or DFNB59), a protein originally linked to a congenital form of hearing loss in Persian families (Delmaghani et al., 2006), localizes to the peroxisomes in hair cells and auditory neurons and mediates a dynamic adaptive reaction involving peroxisomal proliferation/fission to buffer harmful oxidative stress (Delmaghani et al., 2015). Furthermore, the authors conduct a successful proof-of-principle gene therapy approach to correct hearing loss in *Pjvk*-deficient mice.

Over the last decade, a handful of studies linked mutations in *DFNB59* to different forms of autosomal recessive sensorineural hearing loss with variable phenotypic manifestations; however, the function of the protein remained elusive owing to its lack of well-defined functional domains, sorting signals, and limited sequence identity with other proteins. Earlier studies localized *Pjvk* protein and mRNA in afferent auditory neurons and hair cells, but it is also ubiquitously expressed across major mouse organs (Delmaghani et al., 2006; Schwander et al., 2007).

In a systematic phenotypic assessment of *Pjvk*-deficient mice, Delmaghani et al. discovered that young *Pjvk*$^{-/-}$ mice develop a progressive form of sensorineural hearing loss caused by their littermate vocalizations, an otherwise innocuous stimulus for wild-type pups. This exacerbated susceptibility to NIHL under normal acoustic conditions was not due to gross anatomical or histological abnormalities but rather to a progressive postnatal loss of hair cell function and number, suggesting that the auditory phenotype was a result of a specific defect in the cellular adaptation to normal noise levels.

This phenotype correlated with abnormally high levels of oxidative stress markers in the cochlea and gene expression changes consistent with a redox imbalance. Pjvk was shown to localize in peroxisomes, not only in sensory cells of the inner ear, but also in other unrelated cell types, where it mediates peroxisomal proliferation either autonomously or associated with experimentally induced oxidative stress.

Oxidative stress in the inner ear has been long associated with noise-induced damage, including findings of polymorphisms in peroxisomal enzymes and other redox systems linked to NIHL susceptibility (Konings et al., 2007; Wong and Ryan, 2015). In fact, antioxidant therapy has been tested with some success in models of NIHL and age-related hearing loss (Fetoni et al., 2013; Heman-Ackah et al., 2010). The study by Delmaghani et al. is the first one to directly point at peroxisomal dynamics as a first line of antioxidant defense against normal noise exposure, without apparent mitochondrial involvement, and opens many unresolved questions with important implications for the biology of peroxisomes and its relationship with other cellular functions.

One of the major observations of the study is that sound exposure upregulates *Pjvk* expression in hair cells, inducing the proliferation or fission of pre-existing peroxisomes. PJVK is thus proposed to enable an adaptive homeostatic program in response to the accumulation of reactive oxygen species (ROS) induced
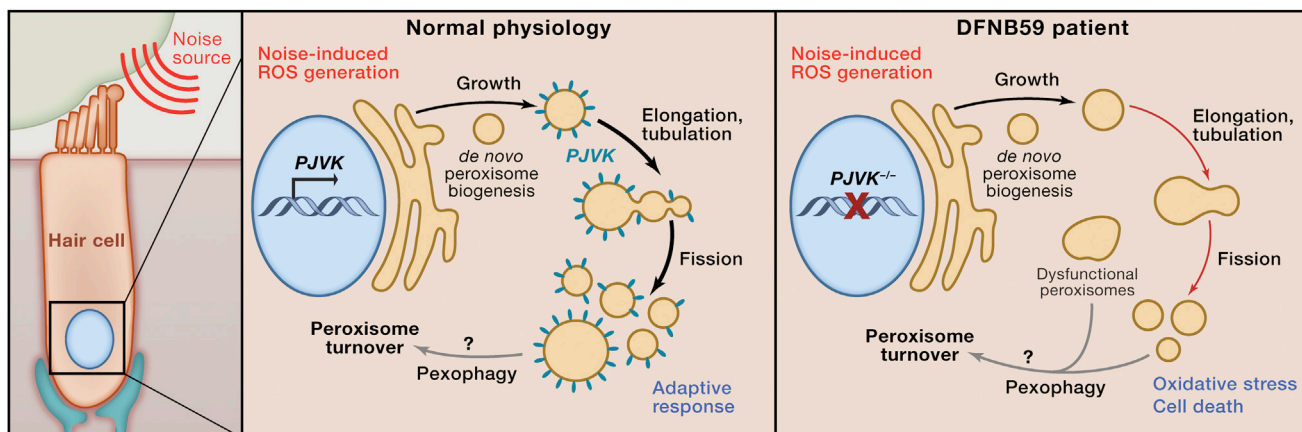
Figure 1. The Biology of PJVK and Peroxisome Dynamics

Sensory cells in the organ of Corti (outer and inner hair cells) carry out mechanotransduction of air pressure waves into electrical signals that are transmitted to primary auditory neurons in the spiral ganglion, which relay these signals for the cortical representation of sound. Hair cells and auditory neurons are particularly sensitive to reactive oxygen species (ROS) generated by exposure to noise. Upregulation of *PJVK* engages an adaptive program modulating peroxisome dynamics, in particular the proliferation/fragmentation of peroxisomes. PJVK locates at the surface of peroxisomes, where it may interact with peroxins and other proteins that mediate crucial steps during the elongation and fission of peroxisomes (PEX11), as well as proteins that regulate pexophagy (PEX3, ATM, PEX5). Thus, PJVK expression contributes to maintain cell viability by increasing the buffering capacity against oxidative stress in part through modulating peroxisome function.

by noise (Figure 1B). Peroxisomes are extraordinarily dynamic ER-derived organelles that interact with other subcellular compartments in the regulation of major cellular functions related to the metabolic and redox status of the cell (Smith and Aitchison, 2013). Peroxisome biogenesis is crucial for normal physiology and its deficiency leads to serious conditions of the Zellweger spectrum, where hearing loss is a common symptom in the context of general neurodegeneration.

Although de novo peroxisome biogenesis is not regulated by PJVK, subtler biogenesis defects secondary to protein trafficking or membrane lipid alterations cannot be excluded. It is also not entirely clear whether the defects observed in peroxisome biology in *Pjvk*$^{-/-}$ hair cells are due to a disruption in the peroxisomal fission machinery or in the turnover of this organelle by pexophagy (Figure 1B). Peroxisomes can operate as a source and a sink of ROS, therefore cellular redox balance depends on a delicate equilibrium between their biogenesis, proliferation, and turnover. A recent report linked the overproduction of ROS to increased pexophagy in a process regulated by the ataxia telangiectasia mutated (ATM) kinase and the peroxisome importer receptor PEX5 (Zhang et al., 2015). This

pathway might also be compromised in *Pjvk*$^{-/-}$ hair cells, leading to the appearance of dysfunctional peroxisomes. The presence of cell-type-specific protein partners may also shed light on the essential function of PJVK in the inner ear.

In addition to NIHL, the most common form of auditory impairment is age-related hearing loss (Wong and Ryan, 2015). Aging is a process characterized by a generalized decline in antioxidant defenses and PJVK might represent a possible therapeutic target on this front, provided that its activity turns out to be sufficient for protecting the sensory epithelium from age-related oxidative damage. Gene therapy approaches using adeno-associated viruses to restore or enhance PJVK function in the cochlea represent an attractive strategy to treat highly prevalent forms of hearing loss in the near future.

REFERENCES

Delmaghani, S., del Castillo, F.J., Michel, V., Leibovici, M., Aghaie, A., Ron, U., Van Laer, L., Ben-Tal, N., Van Camp, G., Weil, D., et al. (2006). Nat. Genet. *38*, 770–778.

Delmaghani, S., Defourny, J., Aghaie, A., Beurg, M., Dulon, D., Thelen, N., Perfettini, I., Zelles, T., Aller, M., Meyer, A., et al. (2015). Cell *163*, this issue, 894–906.

Fetoni, A.R., De Bartolo, P., Eramo, S.L., Rolesi, R., Paciello, F., Bergamini, C., Fato, R., Paludetti, G., Petrosini, L., and Troiani, D. (2013). J. Neurosci. *33*, 4011–4023.

Heman-Ackah, S.E., Juhn, S.K., Huang, T.C., and Wiedmann, T.S. (2010). Otolaryngol. Head Neck Surg. *143*, 429–434.

Konings, A., Van Laer, L., Pawelczyk, M., Carlsson, P.I., Bondeson, M.L., Rajkowska, E., Dudarewicz, A., Vandevelde, A., Fransen, E., Huyghe, J., et al. (2007). Hum. Mol. Genet. *16*, 1872–1883.

Nelson, D.I., Nelson, R.Y., Concha-Barrientos, M., and Fingerhut, M. (2005). Am. J. Ind. Med. *48*, 446–458.

Schwander, M., Sczaniecka, A., Grillet, N., Bailey, J.S., Avenarius, M., Najmabadi, H., Steffy, B.M., Federe, G.C., Lagler, E.A., Banan, R., et al. (2007). J. Neurosci. *27*, 2163–2175.

Smith, J.J., and Aitchison, J.D. (2013). Nat. Rev. Mol. Cell Biol. *14*, 803–817.

Wong, A.C., and Ryan, A.F. (2015). Front. Aging Neurosci. *7*, 58.

Zhang, J., Tripathi, D.N., Jing, J., Alexander, A., Kim, J., Powell, R.T., Dere, R., Tait-Mulder, J., Lee, J.H., Paull, T.T., et al. (2015). Nat. Cell Biol. *17*, 1259–1269.

Histone Marks Direct Chromosome Segregation

Vincenzo Pirrotta[1,*]
[1]Department of Molecular Biology and Biochemistry, Rutgers University, 604 Allison Road, Piscataway, NJ 08854, USA
*Correspondence: pirrotta@dls.rutgers.edu
http://dx.doi.org/10.1016/j.cell.2015.10.043

Germline stem cells divide asymmetrically, producing a self-renewing stem cell and a differentiating progenitor. Xie et al. now show that this depends on two asymmetric events that together partition a genome copy, carrying the old histones to the stem cell daughter and a copy with new, unmarked histones to the differentiating daughter.

Post-translational modifications of the histones around which genomic DNA is wrapped are involved in virtually everything that the genome does, from the various steps of transcription to chromatin replication or DNA damage repair. Are the histone modifications instructive, determining these events? Or are they just facilitators of processes initiated by extracellular signals? A remarkable story from the lab of Xin Chen, the latest installment of which appears in this issue of *Cell* (Xie et al., 2015), sheds more light on this question and shows that histone modifications can also determine how chromosomes are partitioned between daughter cells destined to have different fates during asymmetric cell division.

Chen and co-workers looked at the germline stem cells (GSCs) in the *Drosophila* testes. These stem cells maintain contact with the "hub," a cluster of niche cells that provide important signals for "stemness." The GSCs divide asymmetrically to produce a daughter GSC that remains in contact with the hub and a gonialblast (GB) daughter cell that moves away and initiates differentiation, ultimately to produce spermatocytes. In a remarkable 2012 article, Chen and co-workers reported that parental histone H3 and presumably the entire H3-H4 heterotetramer segregates asymmetrically during GSC division so that the daughter cell that retains stem cell fate inherits the old histone H3 epigenetically marked by GSC-associated post-translational modifications, while the GB daughter cell acquires newly synthesized and unmodified H3 (Tran et al., 2012).

In normal symmetric cell division, the old histones are removed from the DNA just in front of the advancing replication fork and are restored to the DNA on the other side of the fork approximately in equal measure to the two daughter DNA molecules. Newly synthesized histones fill in the gaps resulting from the 2-fold dilution of the old histones. How the redeposition of old histones takes place is poorly understood, but how it is rendered asymmetric in the GSC so that one daughter molecule is the preferred recipient is not understood at all. It may be related to the asymmetry in DNA replication, where one strand is replicated by continuous elongation (leading-strand) while the other (lagging-strand) is replicated discontinuously in short segments. However, this is insufficient to account for the asymmetry since the genome is normally replicated bidirectionally from many origins. Another possibility suggested by the authors is that the two DNA strands are distinguished according to which is the oldest (grandparent) strand. Differential inheritance of the two parental DNA strands has been seen in some cases of asymmetric stem cell division such as mouse muscle stem cells or the *Drosophila* male GSC itself (Evano and Tajbakhsh, 2013; Yadlapalli and Yamashita, 2013) and has been argued to preserve to the stem cell the DNA strand with the fewest replication-induced mutations (the immortal-strand hypothesis, proposed by Cairns, 1975; reviewed by Rando, 2007). No mechanism for the distinction of the two strands is currently known. Regardless of how it is produced, the asymmetric distribution of the old histones suggests that their stem-cell-specific modifications enforce GSC fate while the unmodified histones in the GB daughter leave the cell free to acquire a differentiation program.

In the present paper, however, the authors tackle a second aspect of the asymmetric division. Once the asymmetric old histone distribution is achieved on the replicated DNA molecules, their partitioning needs to be orchestrated such that all the chromosomes with the old histones go to the GSC daughter and all chromosomes with new histones go to the GB daughter cell. To study how this is achieved, Xie et al. used a two-color system previously developed in the Chen lab (Tran et al., 2012) in which a transgene expresses histone H3 labeled with GFP (old histone) but can be switched by a heat shock to express H3 labeled with mKO (new histone). In the mitosis that follows the administration of the heat shock, mKO was found to decorate preferentially the chromosomes that segregate into the GB while the old GFP-H3 was found on the GSC chromosomes. The correct alignment of chromosomes at metaphase is known to involve the Haspin kinase (Dai et al., 2005), which normally phosphorylates the threonine at position 3 of histone H3 in late mitosis. Xie et al. found that, in GSCs, Haspin produces an early phosphorylation event that is specific for the chromatids containing old histones. This earlier event apparently gives the "old" chromosomes the advantage in connecting with the mother centrosome and thus in segregating with the cell that retains GSC identity. The mother centrosome is connected to the surface of the GSC that contacts the hub, and therefore the mitotic spindle is oriented perpendicular to this surface and causes the GSC daughter to remain in contact with the hub while the GB daughter moves away (Yamashita et al., 2007). Signaling from the hub cells is important. For example, hyperactivation of JAK-STAT signaling in the hub cells causes GSCs to divide symmetrically,
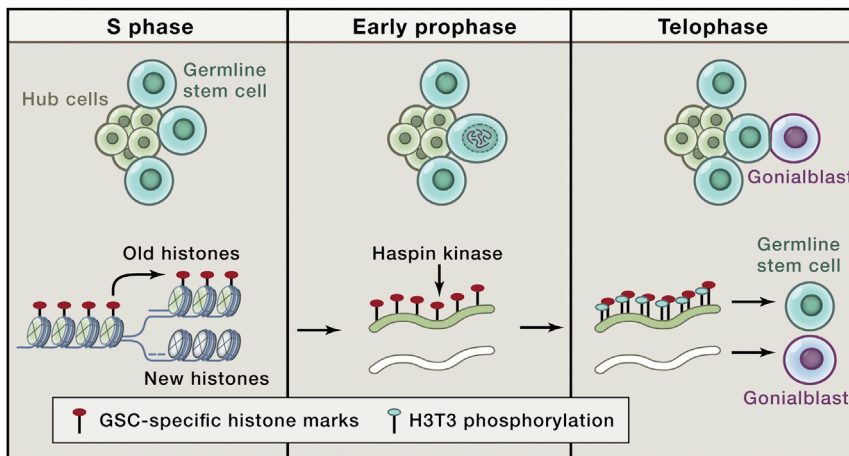
S phase	Early prophase	Telophase

Figure 1. Asymmetric Segregation of Old Histones Is Directed by Phosphorylation of H3T3 and Is Linked to Cell Fate

(Top) Germline stem cells (GSCs) associated with the hub cells divide asymmetrically such that one daughter cell remains a GSC and associated with the hub, while the other becomes a gonialblast (GB), initiates to differentiate, divides symmetrically, and eventually produces spermatocytes. (Bottom) The asymmetric features begin during DNA replication in the GSC. The old histones, presumably with GSC-specific histone modifications, are specifically transferred to one of the daughter DNA molecules (not known which), while newly synthesized histones lacking modifications are deposited on the other DNA molecule. At the beginning of mitosis, an early wave of recruitment of the Haspin kinase phosphorylates histone H3T3 on old histones, which directs attachment to the maternal centrosome. As a result, at telophase, the chromosomes carrying the old histones segregate together to the daughter cell that remains associated with the hub and retains GSC character. The daughter cell with new histones now responds to signals that initiate differentiation.

producing two GSC daughters, each with half of the old histone H3.

To demonstrate the role of H3T3 and its importance for asymmetric segregation, Xie et al. antagonized endogenous H3 by expressing a transgenic H3 in which T3 is mutated to alanine (H3T3A), which cannot be phosphorylated, or to aspartate (H3T3D), which mimics constitutive phosphorylation. H3T3A expression in early germ cells greatly reduces the level of phosphorylated H3T3 in dividing GSCs. As a consequence, the pattern of old histone segregation becomes predominantly symmetric: the dividing GSC can no longer distinguish which chromosomes have the old histones and which have the new. Interestingly, since *Drosophila* has only three large and one very small chromosome, random segregation would occasionally result in the correct asymmetric segregation pattern: all of the "old" chromosomes going with the daughter GSC by chance. With a similar low frequency, the inverse asym-

metry would be also expected: all of the "old" chromosomes going with the GB daughter. Remarkably, this is just what is observed and with a frequency close to that statistically expected. The expression of the H3T3D transgene in early germ cells leads to incorporation of the phospho-mimic H3T3D in both copies of the DNA during replication, and therefore both compete equally in segregation to the daughter cells. Just as with the H3T3A transgene, this results in randomized inheritance of the "old" chromosomes when the GSCs divide (Figure 1).

Even more telling are the consequences of inappropriate segregation of the chromosomes carrying the old histones. In both H3T3A and H3T3D experiments, cells containing a spectrosome, an organelle characteristic of GSCs, decrease in number. In their place, next to the hub region, appear cells containing a fusome, typical of differentiating germ cells, suggesting that GSCs lose their stem cell character and begin to differen-

tiate. In addition, the authors observe heterogeneous "tumors": clusters of cells that replicate actively and exhibit a mixture of GSC-like and GB-like features. This suggests that the daughter cells express their differentiation genes variously in proportion to the number of "new" chromosomes that they received.

To summarize, then, the old histone H3 with its various marks is preferentially transmitted to one genome copy. This instructs Haspin to phosphorylate preferentially the "old" chromosomes, which in turn are preferentially associated with the maternal centrosome. The cell with the "old" chromatin retains stem cell character and stem-cell-specific gene expression, while the cell with "new" chromatin is free to respond to a new program of gene expression leading to differentiation. At least in these cells then, signals from the hub cells are not sufficient to determine "stemness," and the histone modifications are both necessary and sufficient to prevent differentiation. These results raise many fascinating questions. How is Haspin activated for early H3T3 phosphorylation in GSCs? How does it recognize the old histone H3? What "reads" the phosphorylated H3T3? Most intriguing of all, how is the asymmetric distribution of old histones achieved during DNA replication? It is clear that new surprises and insights will be forthcoming as Chen and co-workers continue their study of this process.

REFERENCES

Cairns, J. (1975). Nature *255*, 197–200.

Rando, T.A. (2007). Cell *129*, 1239–1243.

Dai, J., Sultan, S., Taylor, S.S., and Higgins, J.M. (2005). Genes Dev. *19*, 472–488.

Evano, B., and Tajbakhsh, S. (2013). Chromosome Res. *21*, 225–242.

Tran, V., Lim, C., Xie, J., and Chen, X. (2012). Science *338*, 679–682.

Xie, J., Wooten, M., Tran, V., Chen, B.-C., Pozmanter, C., Simbolon, C., Betzig, E., and Chen, X. (2015). Cell *163*, this issue, 920–933.

Yadlapalli, S., and Yamashita, Y.M. (2013). Nature *498*, 251–254.

Yamashita, Y.M., Mahowald, A.P., Perlin, J.R., and Fuller, M.T. (2007). Science *315*, 518–521.

Aire Gets Company for Immune Tolerance

Ludger Klein[1,*]
[1]Institute for Immunology, University of Munich, Goethestr. 31, 80336 Munich, Germany
*Correspondence: ludger.klein@med.uni-muenchen.de
http://dx.doi.org/10.1016/j.cell.2015.10.057

A specialized subset of epithelial cells in the thymus "promiscuously" transcribes thousands of peripheral genes to ensure that developing T cells can test their antigen receptors for dangerous autoreactivity. New findings by Takaba et al. indicate that the transcription factor Fezf2 acts independently of Aire in thymic epithelial cells to generate "genetic noise" for immunological tolerance.

During T cell differentiation in the thymus, every T cell is equipped with a unique antigen receptor generated through somatic rearrangements. Whereas this vast T cell receptor repertoire confers protection against a universe of constantly evolving pathogens, the random nature of the rearrangement process inevitably leads to the emergence of potentially dangerous T cells that recognize the body's own structures. Prior to being released into the blood circulation, immature T cells must therefore "test" their antigen specificity on ligands within the thymic microenvironment and will undergo negative selection against autoreactivity. The scope of self-antigens that are visible to T cells for central tolerance is substantially broadened through the constitutive expression of a plethora of tissue restricted antigens (TRAs) by medullary thymic epithelial cells (mTECs) (Derbinski et al., 2001; Klein et al., 2014). This phenomenon is referred to as "promiscuous gene expression," and it has evolved to facilitate T cell tolerance toward self-antigens that would otherwise be temporally or spatially secluded from the immune system. In this issue of *Cell*, Takaba et al. (2015) identify the transcription factor Fezf2 as a key driver of promiscuous gene expression in mTECs, essential to prevent spontaneous autoimmunity against multiple tissues (Takaba et al., 2015).

Tightly controlled expression of cell-type-specific genetic programs is indispensable for tissue identity and homeostasis, and multiple layers of control act in concert to prevent ectopic gene expression. How do mTECs deliberately violate these rules in order to generate "beneficial genetic noise"? A breakthrough in the field emerged from studies

on the autoimmune polyendocrine syndrome (APS), a monogenically inherited human autoimmune disease resulting from mutations in the autoimmune regulator (Aire) gene. In mice, Aire was found to be essential for "ectopic" TRA expression in mTECs, and knockout of Aire recapitulated several aspects of APS (Anderson et al., 2002). Strikingly, however, it turned out that promiscuous gene expression was not fully abolished in the absence of Aire (Derbinski et al., 2005), suggesting the existence of additional transcriptional regulators. Still, most research in the field focused on deciphering the molecular workings of Aire (Mathis and Benoist, 2009; Peterson et al., 2008), while largely ignoring how Aire-independent promiscuous gene expression is regulated. Here, Takaba et al. (2015) hypothesize that any Aire-independent additional regulators of promiscuous gene expression should be differentially expressed between mTECs and their nearest neighbor, the cortical thymic epithelial cells (cTECs), who share a common precursor with mTECs but do not express TRAs. Among the genes that were differentially expressed in mTECs versus cTECs, they find the transcriptional regulator Fezf2 (forebrain expressed zinc finger 2) to be expressed in mTECs, but not in other thymic stromal cell types. Previous work has implicated Fezf2 in corticospinal motor neuron differentiation, and Fezf2-deficient mice do not survive beyond weaning, explaining why a potential role of Fezf2 in the immune system may have gone unnoticed. By comparing gene expression profiles, Takaba et al. (2015) demonstrate that numerous TRA transcripts are downregulated in Fezf2-deficient mTECs. Importantly, the majority of these Fezf2-dependent TRAs are

not affected in Aire-deficient mTECs, and most Aire-dependent TRAs remain expressed in the absence of Fezf2, strongly supporting an Aire-independent and non-redundant role of Fezf2 in the promotion of promiscuous gene expression. Along these lines, the authors go on to show that Fezf2 deficiency in thymic epithelium elicits a spectrum of autoimmune manifestations that is partly distinct from those seen in Aire-deficient mice.

Remarkably, several aspects of Fezf2's role in the thymus seem to follow a fundamentally different biological and mechanistic logic as compared to what is known about Aire (Figure 1). First, Aire's function as a "transcriptional randomizer" is likely to have evolved simultaneous to, and as a consequence of, the emergence of adaptive T-cell-mediated immunity in jawed vertebrates some 500 million years ago (Saltis et al., 2008). By contrast, Fezf2's evolutionary conserved primary function seems to be that of a master regulator of cell fate specification in corticospinal motor neurons. Thus, Aire appears to be a genuine "tolerance gene," whereas Fezf2 instead might exemplify the evolutionary co-optation of a neuronal gene in a distinct cellular and functional context. Second, Aire lacks a DNA-binding domain and seems to seek out its targets through binding inactive chromatin marks prior to recruiting factors that facilitate "illegitimate" transcription by generating double-strand breaks, fostering mRNA processing, and releasing stalled RNA Polymerase (Mathis and Benoist, 2009; Peterson et al., 2008). By contrast, Fezf2 is a "bona fide" transcription factor that directly binds to target DNA. In neuronal progenitors, Fezf2 binds in the vicinity of the transcriptional start site of more than 10,000 genes (Lodato et al.,
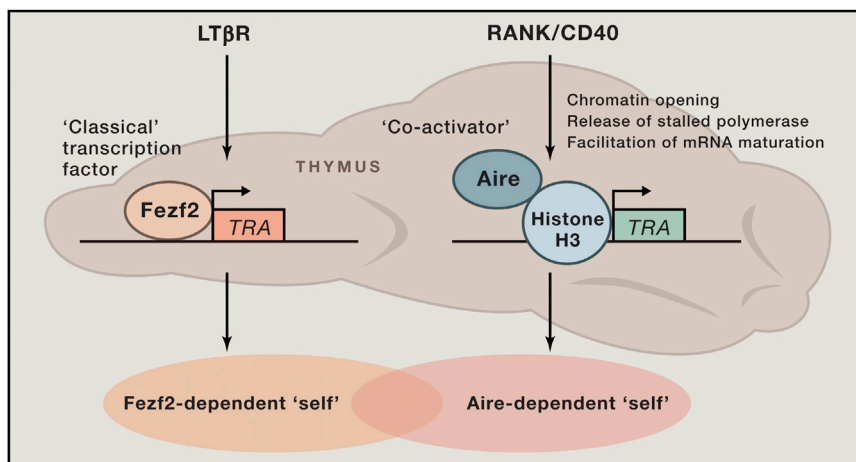
Figure 1. Fezf2 and Aire Promote Promiscuous Gene Expression through Distinct Pathways
Takaba et al. (2015) show that Fezf2 is induced downstream of lymphotoxin β receptor (LTβR) stimulation in mTECs and generates "transcriptional noise" by directly binding to its target genes. By contrast, previous work has indicated that Aire is activated by RANK/CD40 signaling. Aire binds to inactive chromatin marks (unmethylated H3K4) in the vicinity of tissue restricted antigen (TRA) genes and indirectly promotes promiscuous gene expression through recruitment of factors that promote opening of chromatin, release of stalled RNA polymerase, and facilitation of mRNA maturation. Fezf2- and Aire-induced TRAs appear to be largely non-overlapping, resulting in a non-redundant and complementary role of both factors in central T cell tolerance.

2014), which has obvious implications for the spectrum of genes that might be controlled by Fezf2 in mTECs. Finally, whereas Aire expression in mTECs is regulated by the TNF superfamily members RANK and CD40, Takaba et al. (2015) show that the expression of Fezf2 critically depends upon the lymphotoxin (LT) β signaling axis.

Taken together, the work by Takaba et al. (2015) represents a major step forward in our understanding of how promiscuous gene expression in mTECs is

brought about and how it safeguards against autoimmunity. At the same time, these exciting new insights raise a number of questions: do the altered 3D organization of the thymic medulla and subtle alterations in the ratio of mTEC subsets in *Fezf2*$^{-/-}$ mice contribute to faulty T cell selection independent of Fezf2's role in promiscuous gene expression? Given the perplexing observation that individual mTECs only express a subset of Aire-dependent TRAs, does the same apply to Fezf2-dependent transcripts? Does

Fezf2 similarly promote promiscuous gene expression in the human thymus, and if so, are mutations or allelic variants of Fezf2 associated with human autoimmune diseases? And finally, does the complementary action of Aire and Fezf2 account for the full extent of promiscuous gene expression in mTECs, or are as-yet-unknown, additional, and independent factors involved?

REFERENCES

Anderson, M.S., Venanzi, E.S., Klein, L., Chen, Z., Berzins, S.P., Turley, S.J., von Boehmer, H., Bronson, R., Dierich, A., Benoist, C., and Mathis, D. (2002). Science *298*, 1395–1401.

Derbinski, J., Schulte, A., Kyewski, B., and Klein, L. (2001). Nat. Immunol. *2*, 1032–1039.

Derbinski, J., Gäbler, J., Brors, B., Tierling, S., Jonnakuty, S., Hergenhahn, M., Peltonen, L., Walter, J., and Kyewski, B. (2005). J. Exp. Med. *202*, 33–45.

Klein, L., Kyewski, B., Allen, P.M., and Hogquist, K.A. (2014). Nat. Rev. Immunol. *14*, 377–391.

Lodato, S., Molyneaux, B.J., Zuccaro, E., Goff, L.A., Chen, H.H., Yuan, W., Meleski, A., Takahashi, E., Mahony, S., Rinn, J.L., et al. (2014). Nat. Neurosci. *17*, 1046–1054.

Mathis, D., and Benoist, C. (2009). Annu. Rev. Immunol. *27*, 287–312.

Peterson, P., Org, T., and Rebane, A. (2008). Nat. Rev. Immunol. *8*, 948–957.

Saltis, M., Criscitiello, M.F., Ohta, Y., Keefe, M., Trede, N.S., Goitsuka, R., and Flajnik, M.F. (2008). Immunogenetics *60*, 105–114.

Takaba, H., Morishita, Y., Tomofuji, Y., Danks, L., Nitta, T., Komatsu, N., Kodama, T., and Takayanagi, H. (2015). Cell *163*, this issue, 975–987.

Modifications on Translation Initiation

Sarah F. Mitchell[1] and Roy Parker[1,2,*]
[1]Department of Biochemistry and Chemistry, University of Colorado, Boulder, CO 80309, USA
[2]Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA
*Correspondence: roy.parker@colorado.edu
http://dx.doi.org/10.1016/j.cell.2015.10.056

Two studies by Meyer et al. and Wang et al. demonstrate a role for m^{6}A modification of mRNA in stimulating translation initiation. These findings add to the growing number of diverse mechanisms for translation initiation in eukaryotes.

The control of translation initiation is a critical aspect of modulating protein production, particularly when rapid responses to extracellular cues are required, such as during neuronal stimulation or stress conditions (Sonenberg and Hinnebusch, 2007). Translation initiation requires the delivery of the small 40S ribosomal subunit to the mRNA. In eukaryotes, this is primarily achieved in a mechanism that begins with binding of the 5′ mRNA cap by the eIF4F complex, which recruits the 40S subunit pre-bound to a multifactor complex, including eIF3, eIF2, and the initiator tRNA (Figure 1A). The ribosome then scans along the 5′ UTR to the AUG start codon, followed by joining of the large ribosomal subunit, producing a translation competent complex. In a second mechanism, specific mRNA structures referred to as internal ribosome entry sites (IRES) can recruit the 40S subunit either by binding to one of the initiation factors, which then recruits the 40S subunit, or by direct interaction with the 40S subunit, as in the case of CrPV IRES (Figure 1B) (Sonenberg and Hinnebusch, 2007). Two papers in this issue of *Cell* (Meyer et al., 2015; Wang et al., 2015) and a third study (Zhou et al., 2015) now argue that m^{6}A modifications in mRNA can promote translation initiation and suggest two possible mechanisms by which such RNA modifications can lead to ribosome recruitment (Figures 1C and 1D).

Convincing evidence that m^{6}A modifications can stimulate translation comes from the observations that uncapped m^{6}A-containing mRNAs are much more efficiently translated in cell-free extracts than unmodified mRNAs, and m^{6}A-modified mRNAs assemble translation initiation complexes in reconstituted systems

in the absence of the eIF4F complex, unlike unmodified mRNAs (Meyer et al., 2015). Strikingly, a single m^{6}A in the 5′ UTR is sufficient to boost cap-independent translation both in extracts and when mRNAs are introduced into cells by transfection (Meyer et al., 2015; Zhou et al., 2015). Evidence that m^{6}A modifications promote translation in vivo is that depletion of the METTL3 m^{6}A methyl-transferase reduces ribosome occupancy for mRNAs with 5′ UTR m^{6}A modification sites (Meyer et al., 2015), and on mRNAs that are bound by YTHDF1, an m^{6}A-binding protein (Wang et al., 2015). Moreover, for the Hsp70 mRNA, the extent of m^{6}A modification corresponds to the rate of protein production and polysome occupancy during heat shock (Meyer et al., 2015; Zhou et al., 2015). Finally, transfected mRNAs with a cap unable to stimulate translation are effectively translated under stress conditions if they contain a 5′ UTR m^{6}A modification (Zhou et al., 2015).

Meyer et al. (2015) provide three observations that m^{6}A stimulates cap-independent translation through interactions with eIF3, thereby leading to ribosome recruitment (Figure 1C). First, in a reconstituted system, eIF3 preferentially cross-links to RNA with m^{6}A modifications. Second, in vivo, eIF3-binding sites defined by cross-linking significantly overlap with m^{6}A modification sites in 5′ UTRs. Third, overexpression of the FTO demethylating enzyme reduces the association of 5′ UTR m^{6}A-modified mRNAs with eIF3. Interestingly, the authors demonstrate that eIF3 prefers to bind m^{6}A-modified mRNA when the modification is within the expected GAC sequence context. This may correlate with the observation that m^{6}A is not able to stimulate translation in

all 5′ UTRs, demonstrating the importance of context (Zhou et al., 2015). However, whether this observation is due to differences in eIF3 interactions has not been determined.

In contrast, several observations lead Wang et al. (2015) to suggest that m^{6}A modifications in the 3′ UTR, and possibly the coding region, may enhance translation by binding the C-terminal domain of the YTHDF1 m^{6}A-binding protein, which then recruits the translation initiation complex through its N-terminal domain (Figure 1D). First, knockdown of YTHDF1 leads to reduced ribosome occupancy on mRNAs bound by YTHDF1. Second, tethering the N-terminal domain of YTHDF1 to an mRNA leads to some increase in translation. Finally, YTHDF1 co-purifies with a large number of proteins, including eIF3 in a RNase-resistant manner, suggesting that the interaction with eIF3 allows YTHDF1 to promote translation of m^{6}A modified mRNA (Wang et al., 2015). Interestingly, Meyer et al. (2015) do not see changes in translation profiles in YTHDF1 knockdown cells when examining 5′ UTR, 3′ UTR, or all m^{6}A-modified mRNAs, suggesting that YTHDF1 effect on translation would be limited to a subset of m^{6}A-modified mRNAs.

A number of questions remain. Do these two proposed mechanisms for m^{6}A stimulation of translation cooperate or compete in different contexts? How does the growing number of m^{6}A-binding proteins (YTHDF1, YTHDF2, eIF3, etc.) recognize specific binding sites? eIF3 interacts preferentially with m^{6}A modifications found in the 5′ UTR, but these are a minority of such modifications in the transcriptome. What other protein factors or local mRNA features define an
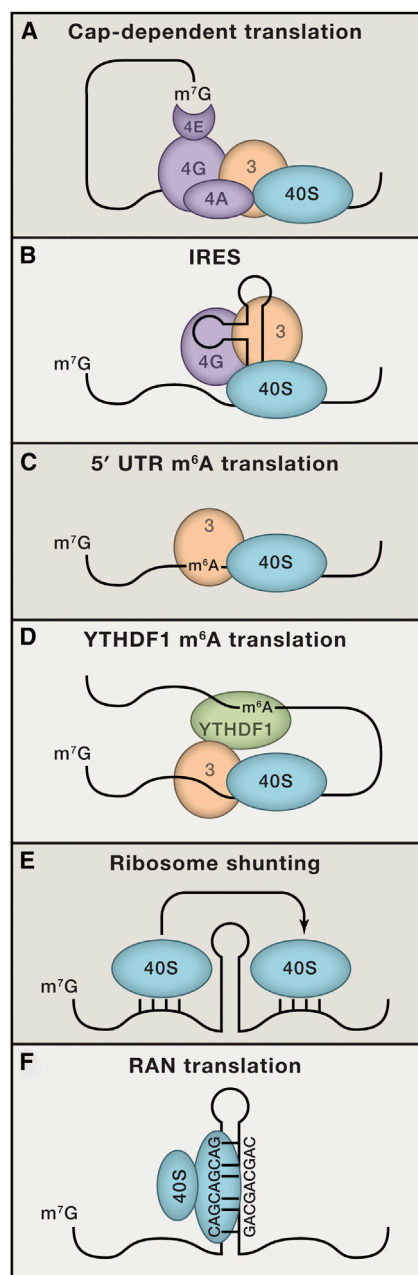
Figure 1. Mechanisms of Translation Initiation in Eukaryotes

(A) Cap-dependent translation initiation. eIF4F complex binds the 5′ cap of mRNA and then recruits the 40S ribosomal subunit pre-bound to a multifactor complex, including eIF3, eIF2, and the initiator tRNA, to start translation initiation.

(B) IRES-stimulated translation initiation. Some mRNAs contain specific IRES structures that recruit the 40S subunit either indirectly by binding to one of the initiation factors.

(C) 5′ UTR m^6A-mediated translation initiation. Translation initiation is stimulated by m^6A modification of mRNA 5′ UTR via direct recruitment of eIF3.

(D) YTHDF1-mediated translation initiation. Translation initiation is stimulated by m^6A modification of the 3′ UTR of mRNA through recruitment of

YTHDF1, which subsequently recruits the translation initiation complex.

(E) Ribosome shunting. Ribosomal RNA base pairs with mRNA leading to the translocation of 40S subunit from the cap region to internal start codons for initiation.

(F) Repeat-associated non-AUG (RAN) translation. Translation initiation can occur at disease-associated CAG repeats.

eIF3-binding site to prevent binding in other regions of the mRNA? How is the competition between m^6A-binding factors properly balanced? Finally, since methylation is reversible, like many chromatin modifications, it will be important to determine the mechanisms regulating the rates of methylation and demethylation of specific sites.

A broader point from these papers is that eukaryotic cells contain a growing diversity of mechanisms for translation initiation, which has implications for our understanding of the predicted proteome. In addition to canonical cap-dependent translation, IRES, and now m^6A modification-stimulated initiation, other mechanisms exist (Figures 1E and 1F). For example, ribosome shunting involves the translocation of 40S ribosomes from the cap region to internal sites for initiation, which can lead to the use of internal AUGs, and/or the skipping of 5′ UTR RNA structures that would otherwise inhibit translation (Figure 1E) (Chappell et al., 2006). A mechanism by which ribosomes might be recruited to mRNAs independent of the cap is suggested by the binding of eIF3 to stem loops in specific 5′ UTRs, which can result in either stimulation or inhibition of translation (Lee et al., 2015). Ribosome profiling studies have also identified translation initiation sites at near-cognate start codons, suggesting that the start site, as well as the initiation complex, is malleable (de Klerk and 't Hoen, 2015). A striking example of an unexpected mode of translation is seen in the case of repeat-associated non-AUG (RAN) translation, which occurs at disease-associated CAG repeats (Figure 1F) (Zu et al., 2011). Although the mechanism of RAN translation is not known, a reasonable prediction is that cells use that same non-AUG-dependent mode of translation in some context.

One has to anticipate that cells use additional yet-to-be-discovered mecha-

nisms to recruit ribosomes to mRNAs. For example, is has been suggested that some mRNAs recruit eukaryotic ribosomes by direct base pairing to rRNAs, similar to the bacterial mechanism of initiation in which the Shine-Dalgarno sequence 5′ of the start codon base pairs to the small ribosomal subunit (Deforges et al., 2015). Moreover, one speculates that evolution is likely to have chanced upon sequence-specific RNA-binding proteins that interact with eIF3 or other initiation factors to recruit the 40S subunit in a cap-independent manner. Finally, it remains possible that other mRNA base modifications will also stimulate translation initiation in some context.

The growing diversity of translation initiation mechanisms allows the cell to preferentially control the translating population of mRNAs under different conditions. For example, cap-dependent translation is inhibited when the TOR pathway is inactive, such as under nutrient deprivation or stress. However, to survive such conditions, the cell must produce stress-response proteins, which can be done by utilizing cap-independent mechanisms of initiation. Consistent with this view, Meyer et al. (2015) observe that Hsp70 translation is stimulated via m^6A during heat shock, when cap-dependent translation is inhibited. They also analyze m^6A modification across the genome under heat and UV stress and find that m^6A modifications specifically increase in the 5′ UTR during stress. The increase in m^6A 5′ UTR modifications during heat shock may be due to nuclear import of YTHDF2 during heat stress, which allows it to compete with the demethylase FTO (Zhou et al., 2015). Importantly, many known variations of translation initiation have been identified under conditions considered non-standard, such as during development or under stress. As shown by Meyer et al. (2015) for m^6A modifications, these variations of translation initiation may be functional during normal growth conditions but are likely more active during conditions in which inhibition of cap-dependent translation allows alternative mechanisms to be more competitive. Thus, studies of translation mechanisms in non-traditional cellular conditions may reveal an even broader set of translation initiation mechanisms.

REFERENCES

Chappell, S.A., Dresios, J., Edelman, G.M., and Mauro, V.P. (2006). Proc. Natl. Acad. Sci. USA *103*, 9488–9493.

de Klerk, E., and 't Hoen, P.A. (2015). Trends Genet. *31*, 128–139.

Deforges, J., Locker, N., and Sargueil, B. (2015). Biochimie *114*, 48–57.

Lee, A.S.Y., Kranzusch, P.J., and Cate, J.H.D. (2015). Nature *522*, 111–114.

Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S.-B., and Jaffrey, S.R. (2015). Cell *163*, this issue, 999–1010.

Sonenberg, N., and Hinnebusch, A.G. (2007). Mol. Cell *28*, 721–729.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). Cell *161*, 1388–1399.

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R., and Qian, S.-B. (2015). Nature *526*, 591–594.

Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A.C., et al. (2011). Proc. Natl. Acad. Sci. USA *108*, 260–265.

Primer

Design and Analysis of Single-Cell Sequencing Experiments

Dominic Grün[1,2,3] and Alexander van Oudenaarden[1,2,*]
[1]Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences), 3584 CT Utrecht, the Netherlands
[2]University Medical Center Utrecht, Cancer Genomics Netherlands, 3584 CX Utrecht, the Netherlands
[3]Max Planck Institute of Immunobiology and Epigenetics, D-79108 Freiburg, Germany
*Correspondence: a.vanoudenaarden@hubrecht.eu
http://dx.doi.org/10.1016/j.cell.2015.10.039

Recent advances in single-cell sequencing hold great potential for exploring biological systems with unprecedented resolution. Sequencing the genome of individual cells can reveal somatic mutations and allows the investigation of clonal dynamics. Single-cell transcriptome sequencing can elucidate the cell type composition of a sample. However, single-cell sequencing comes with major technical challenges and yields complex data output. In this Primer, we provide an overview of available methods and discuss experimental design and single-cell data analysis. We hope that these guidelines will enable a growing number of researchers to leverage the power of single-cell sequencing.

Introduction

Understanding the development and function of an organ requires the knowledge of its constituents, i.e., of all the different cell types the organ is composed of. It is still common practice to distinguish cell types based on a small set of marker genes. These can be used to isolate sub-populations of cells, e.g., by fluorescence-activated cell sorting (FACS), which can then be characterized by population-based assays such as next-generation sequencing. This approach is inherently constrained, since a pre-selection of marker genes limits the resolution and variability within a marker-gene-expressing sub-population of cells cannot be resolved. Moreover, even cells of the same type can show substantial gene expression variability leading to phenotypic variation (Eldar and Elowitz, 2010; Munsky et al., 2012; Snijder and Pelkmans, 2011). The ideal approach to profile the cell type composition of an organ or to explore transcriptome heterogeneity across cells of the same type is a separate analysis of individual cells randomly drawn from a sample. Single-cell analysis of a small number of genes can be performed with imaging-based methods such as single-molecule fluorescence in situ hybridization (Raj et al., 2008) or by flow cytometry, exploiting cell surface markers or fluorescent reporter proteins. Single-cell transcriptome analysis, on the other hand, is an experimental approach to obtain an unbiased view of all mRNAs present in a cell. Already by 1992 the expression of selected genes in individual neurons had been quantified by Southern blotting after amplifying the entire pool of mRNAs from a cell (Eberwine et al., 1992). Single-cell transcriptome sequencing had initially been applied by the Surani laboratory in 2009 (Tang et al., 2009). Over the last five years, a number of single-cell mRNA-sequencing methods with improved sensitivity and reduced technical noise have been introduced (Hashimshony et al., 2012; Islam et al., 2011, 2014; Picelli et al., 2013; Ramsköld et al., 2012; Sasagawa et al., 2013). These methods have been used to discriminate cell types in healthy tissues (Jaitin et al.,

2014; Zeisel et al., 2015), to study differentiation dynamics (Treutlein et al., 2014), to discover rare cell types (Grün et al., 2015), to investigate the transcriptome response upon external signals (Shalek et al., 2013, 2014), or to profile tumor heterogeneity (Patel et al., 2014).

The genotypic variation that underlies cell-to-cell differences can be explored by single-cell genomics. In a landmark study, sequencing of the genomic DNA from single-tumor cell nuclei was employed to profile chromosome copy numbers in order to elucidate clonal expansion and tumor evolution (Navin et al., 2011). Subsequently, a number of improved methods have been published permitting the detection of genomic copy number variations and other structural rearrangements with increasing spatial resolution (Falconer et al., 2012; Gole et al., 2013; Wang et al., 2012; Zong et al., 2012).

In this Primer, we give an overview of the available techniques for genome and transcriptome sequencing, discuss the specific aspects and limitations of each method, and propose guidelines for designing single-cell sequencing experiments. Since any single-cell sequencing technique is based on amplification of minute amounts of material leading to substantial technical noise (Brennecke et al., 2013; Grün et al., 2014), data processing and analysis require extra care. We will discuss in depth all necessary steps for data acquisition, filtering, and analysis, with a focus on single-cell transcriptomics.

Isolating Single Cells for Sequencing

To perform any kind of single-cell sequencing assay, individual cells first have to be isolated from the system of interest. The method of choice to purify thousands of single cells is FACS. With unrestricted sorting gates, random samples of cells can be purified. Alternatively, sorting gates can be set based on scatter properties reflecting the morphology and composition of a cell. Fluorescently labeled antibodies against cell surface markers provide another strategy to purify sub-groups of cells.

Current technology permits the simultaneous measurement of up to 20 parameters per cell and thus highly specific sub-groups of cells can be isolated by FACS (Chattopadhyay and Roederer, 2012). These can be sorted directly into 96- or 384-well plates amenable to subsequent single-cell sequencing. Importantly, the parameter information can be allocated to each well. However, flow cytometry requires a large starting volume, and sorting errors can lead to wells with cell doublets or empty wells.

Micromanipulation provides an alternative approach when only a few cells are available and visual inspection of a cell is desired prior to sequencing. Here, cells are aspirated with a glass micropipette under the microscope. However, this method is very laborious and not well suited to high-throughput single-cell analysis.

More recently, microfluidic devices became available that enable sorting single cells into individual compartments where cells can be visually monitored and further processed. This Fluidigm C1 autoprep system is particularly suited to single-cell sequencing (Islam et al., 2014; Pollen et al., 2014). A shortcoming of this method is the fixed chip architecture that limits the selection of cells to a certain size window. A more detailed discussion of single-cell isolation methods has been published recently (Saliba et al., 2014).

Comparison of Whole-Genome Amplification Techniques

Being able to sequence the genome of individual cells permits the investigation of many relevant questions. Over the lifetime of an organism, cells undergo multiple rounds of division. During each cell division, DNA replication errors can escape the DNA repair machinery with a small probability and can lead to so-called somatic mutations, which can give rise to cancer (Alexandrov and Stratton, 2014) and other diseases (Biesecker and Spinner, 2013). Moreover, a surprisingly high frequency of chromosomal abnormalities has been observed during mammalian germline development (Nagaoka et al., 2012). All types of germline and somatic genome mutations, comprising substitutions, insertions and deletions (indels), copy number variations (CNV) and structural rearrangements, can in principle be detected by DNA sequencing. Moreover, genetic inheritance can be studied by quantifying maternal and paternal allele frequencies based on single-nucleotide polymorphisms (SNPs). However, a single mammalian cell contains less than 10 pg of DNA, necessitating whole-genome amplification (WGA) prior to sequencing or microarray-based analysis. Currently available WGA principles are based on polymerase chain reaction (PCR), multiple displacement amplification (MDA), or a combination of the two. PCR-based strategies initiate amplification by either priming with random oligonucleotides (Cheung and Nelson, 1996; Zhang et al., 1992) or by universal adaptors that are ligated to DNA fragments after enzymatic digestion (Klein et al., 1999). MDA utilizes isothermal amplification by a DNA polymerase with strand displacement activity, typically ϕ29, initiated by random priming of denatured DNA (Dean et al., 2002). The polymerase possesses high processivity and generates DNA amplicons up to 10 kb in length. Upon contact between the 3′ end of an amplicon and the 5′ end of an adjoining amplification product during synthesis, the latter gets displaced, liberating the strand for further amplifi-

cation. All available methods introduce technical artifacts originating from non-uniform genome coverage, in particular due to biased amplification of sequence rich in cytosine and guanosine (GC-bias), preferential allelic amplification or allelic dropout, base copy errors, and chimeric DNA molecules (Macaulay and Voet, 2014). Since the prevalence of a particular type of error depends on the method, the experimental technique should be selected based on the desired readout. In general, random primed PCR-based methods achieve a highly uniform amplification but yield only sparse coverage of the genome and are therefore well suited for low-resolution copy number variant detection down to a length scale of 60 kb (Möhlendick et al., 2013). Due to the high processivity in combination with the strand displacement activity, a much better genome coverage can be achieved with MDA. Together with the high fidelity of the ϕ29 polymerase, this method is better suited for SNP calling. On the other hand, MDA yields highly non-uniform amplification, and the observed biases are only partially explained by the GC bias. This implies the risk of false positives if MDA is used for CNV detection. Moreover, both PCR- and MDA-based techniques produce chimeric DNA molecules, introducing artifacts that can be interpreted as indels or structural rearrangements (Voet et al., 2013).

A technique for obtaining broad coverage of the genome together with uniform amplification was recently developed that combines pre-amplification by a polymerase with strand-displacement activity and amplification by PCR (Zong et al., 2012). The method, termed multiple annealing and looping-based amplification cycles (MALBAC), pre-amplifies DNA with a strand-displacement polymerase and generates amplicons with complementary ends. This complementarity induces loop formation and prevents the amplicon from being used as a template during subsequent cycles to attain close-to-linear amplification. After five cycles of pre-amplification, the material is amplified exponentially by PCR. Sequencing of MALBAC-amplified material from a single cell yielded 93% genome coverage at an average 25× sequencing depth. Due to the improved uniformity and a substantially lower allele dropout rate in comparison to MDA (~1% for MALBAC versus ~31%–65% for MDA [Leung et al., 2015; Zong et al., 2012]), MALBAC shows higher detection efficiency for SNPs and CNVs. The residual false-positive rate of MALBAC (~4×10^{-5}) is due to the relatively low fidelity of the polymerase and could be reduced by sequencing two or three daughter cells derived from the same mother cell. MALBAC is therefore well suited for the simultaneous characterization of SNPs and CNVs.

Another strategy to eliminate amplification biases and alleviate non-uniformity of genome coverage inherent to MDA is the reduction of the reaction volume, for instance, by using nano-liter reaction wells (Gole et al., 2013). This method, termed micro-well displacement amplification system (MIDAS), reduces reaction volume by ~1,000-fold in comparison to conventional MDA, thereby increasing the effective template concentration and reducing contamination. Traditional whole-genome amplification requires extensive purification in order to reduce environmental contamination. In another study, a nano-liter reaction volume was obtained by applying microfluidics to WGA, thereby minimizing amplification error and yielding an extremely low error rate of 4×10^{-9} (Wang et al., 2012). A more detailed comparison

of the available WGA methods has been presented elsewhere (Macaulay and Voet, 2014).

Following WGA, quantification can be performed either by DNA microarrays or by next-generation sequencing. Microarrays can resolve larger CNVs, down to less than 100 kb (Möhlendick et al., 2013), and SNP arrays have been used to infer genome-wide haplotypes from a single human cell with high accuracy (Fan et al., 2011). Moreover, family-based phasing approaches were successfully applied for haplotyping human embryos (Otto-lini et al., 2015). Next-generation sequencing offers the advantage that every amplified base of the DNA is quantified with digital precision and thus enables detection of all types of anomalies, while the microarray readout is constrained by the probe library. Moreover, paired-end sequencing provides additional information since the mapped loci of the two ends together with the fragment size distribution can reveal structural rearrangements within the genome. Of note, sequencing the genome of a single cell with the Strand-seq protocol retains the strand information and allows the derivation of sister chromatid exchange (Falconer et al., 2012). This method provides valuable information for de novo genome assembly or the revision of existing assemblies.

Although substantial progress has been made toward attaining high coverage and uniformity of WGA, there is room for improvement of existing methods, as recently demonstrated by the development of MALBAC (Zong et al., 2012) or by scaling down the reaction volume in order to reduce amplification (Gole et al., 2013; Wang et al., 2012).

Analysis of Single-Cell Genome Sequencing Data

The first step in the data analysis after obtaining a file with sequencing reads is mapping to a reference genome. The genomic DNA sequence for most model organisms can be readily obtained from various online databases, such as the UCSC genome browser (Meyer et al., 2013) or www.ensembl.org (Cunningham et al., 2015). Prior to mapping, it is advisable to inspect the read quality and trim low-quality bases as well as remaining adaptor sequences at the end of the reads. However, if the remaining read length is too short, reads should be discarded in order to avoid erroneous mappings. Furthermore, it is recommended to remove PCR duplicates. After the mapping is performed, reads that map to more than a single locus should be discarded or counted with reduced uniform weight for each locus, such that the weights of each read add up to one. Subsequent processing depends on the type of analysis. To determine CNVs, local variability in read coverage can be alleviated by segmenting the genome into bins. After correcting the number of reads within each bin for GC bias CNV breakpoints can be determined based on a comparison of the change in read number between adjacent bins to a background model (Venkatraman and Olshen, 2007; Zhang et al., 2013). For instance, the circular binary segmentation algorithm (Venkatraman and Olshen, 2007) uses t-statistics with a permutation reference distribution to infer p values for breakpoints. Another study employed a hidden Markov model for CNV detection, with the hidden states corresponding to the local copy number (Zong et al., 2012). Abnormal copy numbers in a cancer cell were inferred after eliminating the amplification bias with a normalization factor derived from a non-

cancer cell. The emission probabilities of this model correspond to binary vectors indicating whether the cancer cell had higher copy number than the normal cell. The numerous published methods for CNV detection using next-generation sequencing were discussed in a recent review (Zhao et al., 2013).

The genome analysis toolkit GATK comprises a bundle of methods for processing of next-generation sequencing data and variant calling (McKenna et al., 2010). In particular, it contains a Bayesian framework that can be used for SNP detection. For each locus, the genotype with the highest posterior probability is emitted if its log odds ratio exceeds a defined threshold. A comprehensive overview and a comparative analysis of existing software tools for SNP calling from next-generation sequencing data can be found in the literature (Nielsen et al., 2011). An advanced method for the detection of structural rearrangements utilizes paired-end read information by creating a bona fide list of discordantly mapped read pairs and identifies candidate rearrangements supported by more than one pair from this list (Voet et al., 2013).

Although correction of GC bias is possible (Baslan et al., 2012; Voet et al., 2013; Zhang et al., 2013), other confounding factors such as allelic dropout or preferential allelic amplification cannot be easily corrected for and may introduce false positives in SNP and CNV detection. Random sequencing errors represent another source of uncertainty for SNP detection. To increase confidence, repeated detection of a given anomaly in more than a single daughter of the same cell is required (Zong et al., 2012). Finally, another confounding factor can be the cell-cycle phase since replication domains of cells in S phase can be mistaken as genuine structural aberrations (Van der Aa et al., 2013). This problem can be avoided by using only nuclei in G1 or G2/M phase. Limiting the analysis to G2/M phase comes with the additional advantage of having duplicated material after replication of the entire genome (Wang et al., 2014).

Comparison of Single-Cell Transcriptome Sequencing Techniques

Measuring gene expression in populations of cells with microarrays or RNA sequencing masks the true distribution of gene expression levels across cells, and it is therefore crucial to quantify gene expression in individual cells. The major hurdle is to obtain sufficient material from an individual cell that can be sequenced with standard next-generation sequencing protocols. Different methods for the amplification of the sub-picogram amount of mRNA from a single cell have been developed and are discussed in detail below. The main problem with any of these methods is the presence of amplification bias, which can distort the relative abundances of mRNAs from different genes.

In the past, amplified RNA from single cells was quantified with microarrays (Iscove et al., 2002). More recently, a number of single-cell sequencing techniques with improved sensitivity were developed. The first protocol for single-cell sequencing was published in 2009 by the Surani laboratory (Tang et al., 2009) and was subsequently used to trace the derivation of mouse embryonic stem cells from the inner cell mass with single-cell resolution (Tang et al., 2010). The amplification method is based on pull-down and reverse transcription of polyadenylated RNA using a poly(T) primer with a specific anchor sequence. Thereafter, the

single-stranded cDNA is polyadenylated and second-strand synthesis is performed using a poly(T) primer with another anchor sequence. The double-stranded cDNA is then PCR amplified from primers against the two anchor sequences, and the resulting material is fragmented prior to library preparation. Although SOLiD sequencing was applied initially, the protocol is compatible with Illumina sequencing, which has become the prevalent method for single-cell sequencing. An initial method that leveraged the integration of DNA barcodes to allow pooling of the material extracted from different cells along with preservation of strand information was termed single-cell tagged reverse transcription (STRT) (Islam et al., 2011). This technique exploits the template-switching property of the reverse transcriptase to tag the 5′ end of polyadenylated mRNA molecules. Following PCR amplification, the tagged ends are pulled down and sequenced, yielding a strong 5′ end bias of the sequencing read. A complementary method termed cell expression by linear amplification and sequencing (CEL-seq) amplifies polyadenylated mRNA linearly from a T7 promoter introduced during cDNA synthesis, thereby reducing amplification bias and alleviating the need for a template switch. Here, only fragments derived from the 3′ end of the mRNA are sequenced. CEL-seq and STRT-seq integrate a barcode into the sequencing primer, a stretch of eight nucleotides that uniquely labels all mRNAs from the same cell. In order to robustly assign mRNAs to different cells, each pair of barcodes should differ in at least two positions. To obtain read coverage along the entire transcript, the Smart-seq and Smart-seq2 methods are a more recent alternative (Picelli et al., 2013; Ramsköld et al., 2012). Similar to STRT, this approach reverse transcribes polyadenylated RNA and exploits the template-switching capacity of the reverse transcriptase. However, using the Nextera technology, the Tn5 transposase simultaneously fragments the cDNA and ligates sequencing adaptors to all fragments, yielding sequencing reads derived from the entire transcript. Another more recent method that yields read coverage of the entire gene body is the Quartz-seq method, which is similar to the approach developed by the Surani laboratory (Tang et al., 2009) but achieves higher sensitivity and reproducibility (Sasagawa et al., 2013). Moreover, two whole-transcript sequencing methods for low starting material have been published, exploiting either Φ29 DNA polymerase or semi-random-primed PCR based amplification (Pan et al., 2013).

To reduce amplification bias, unique molecular identifiers (UMI) (Kivioja et al., 2012) have been integrated into some of the single-cell sequencing protocols. UMIs are stretches of four to ten random nucleotides integrated into a sequencing primer and serve as a random barcode for each mRNA molecule. Upon binding of the sequencing primer, each mRNA is uniquely labeled with a random barcode and the labeled end of the mRNA is amplified along with the barcode. After sequencing, the amplification bias can be eliminated by counting each label only once instead of the reads derived from all amplicons. The number of UMIs can thus be directly translated into the number of sequenced molecules from a cell after application of a mathematical correction to account for the effect of random counting statistics (Grün et al., 2014; Kivioja et al., 2012).

UMIs can only be used for methods that sequence a single tag derived from a given mRNA and have been integrated, for example, into the STRT protocol (Islam et al., 2014) and into modified versions of CEL-seq (Grün et al., 2014; Jaitin et al., 2014). It has been shown that counting UMIs instead of reads leads to a 2-fold reduction of technical noise (Grün et al., 2014).

An overview of three common single-cell sequencing methods is given in Figure 1. In order to select the appropriate sequencing technology, one has to consider the goal of the experimental study. For example, in order to investigate gene expression heterogeneity between cells, the technical variability should be minimized and a technology that allows integration of UMIs should be chosen. However, if information along the entire transcript is required, for instance, to examine splicing patterns, a technology that yields whole-transcript coverage should be selected. Moreover, methods that sequence either the 5′ or 3′ end of a transcript provide single-cell information on the transcriptional start site or polyadenylation site usage, respectively. Another aspect to consider is ease of the experimental procedure and sequencing cost per cell. An increasing number of protocols can be conveniently performed on the Fluidigm C1 multifluidic auto-prep system. This device permits the isolation and processing of the cells, with the important benefit that each cell is imaged. This allows controlling for multiple cells per well and empty wells. However, sequencing-chips that can be used in this device come in fixed geometries and preferentially select cells of particular sizes. Moreover, this technology is relatively expensive. A massively parallel RNA single-cell sequencing framework termed MARS-seq (Jaitin et al., 2014) has been developed based on the CEL-seq technology and employs automated processing of single cells sorted into 384-well plates.

Recently, two advanced droplet-based microfluidic methods, termed Drop-seq (Macosko et al., 2015) and inDrop sequencing (Klein et al., 2015), were published that can dramatically increase the throughput to thousands of cells and at the same time minimize the sequencing costs. Both of these methods rely on the separation of cells into nanoliter-sized aqueous droplets in an oil-water emulsion, which contains sequencing primers with unique cell barcodes and UMIs. The co-occurrence of multiple cells in the same droplet is avoided by a low cell-loading rate into the droplets. In Drop-seq cDNA is PCR amplified, while inDrop sequencing amplifies cDNA by in vitro transcription akin to CEL-seq. In terms of technical noise and sensitivity, these methods compare favorably to previous protocols. Drop-seq was used to characterize mouse retinal cells, while inDrop sequencing was applied to explore cellular heterogeneity during mouse embryonic stem cell differentiation. However, the set-up for neither of these methods is commercially available, and the user is required to build a microfluidic device based on the information provided by the authors.

Although there has been much progress in increasing throughput and lowering costs of single-cell sequencing, there has been only a moderate improvement of the sequencing sensitivity during the last 3 years. The most common method to quantify sensitivity is the usage of external spike-in RNA of known concentration. The spike-in concentration should be chosen such that spike-in RNA contributes 1%–5% of the number of mRNA molecules (Hashimshony et al., 2012). Most of the recently published sensitivity estimates are derived from a set of 92 spike-in RNAs designed by the External RNA Controls
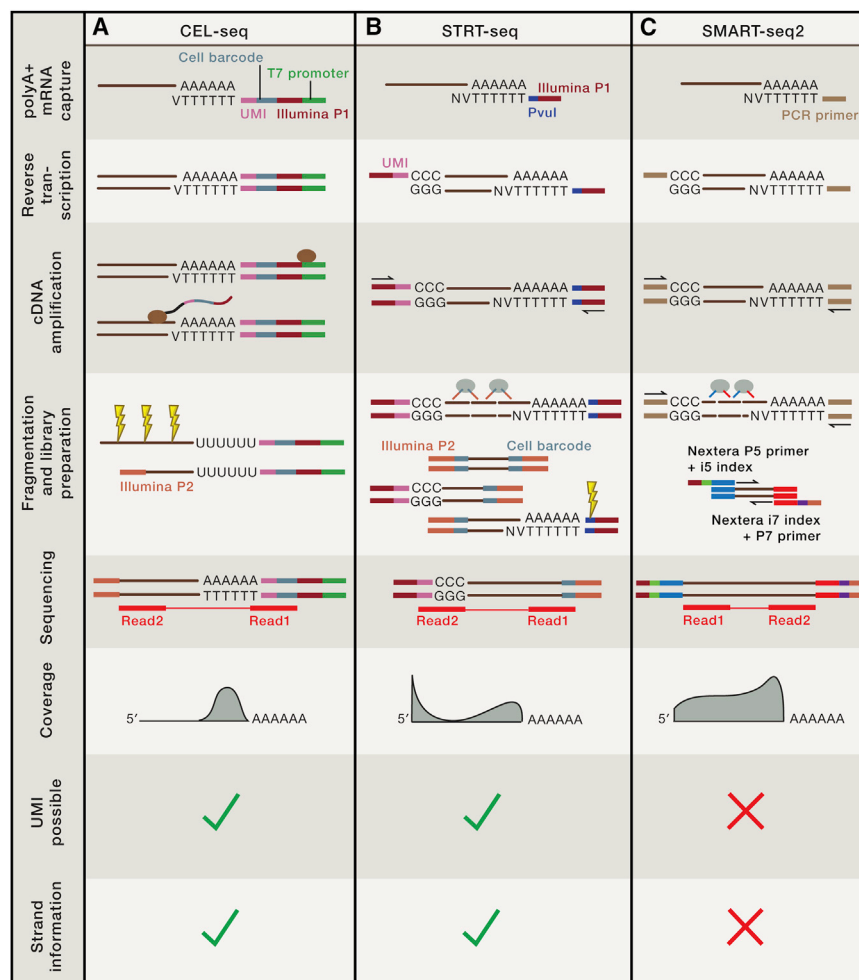
Figure 1. Three Common Experimental Protocols for Single-Cell Sequencing

(A) CEL-seq. Polyadenylated mRNA is reverse transcribed from an Oligo dT primer containing the Illumina P1 adaptor, a cell barcode, and a T7 promoter. The sequencing primer can, in principle, also accommodate a UMI. Following second-strand synthesis, the cDNA is amplified by in vitro transcription from the T7 promoter, and the Illumina P2 adaptor is ligated after fragmentation. The sequencing reads are thus derived from the mRNA 3′ end.

(B) STRT-seq. Polyadenylated RNA is reverse transcribed from an Oligo-dT pimer containing the Illumina P1 adaptor and a PvuI restriction site. After full-length reverse transcription, a template-switching oligo with another Illumina P1 adaptor and the UMI is added to the 5′ end of the transcript. Following second-strand synthesis, the cDNA is then PCR amplified using primers complementary to the Illumina P1 adaptor. Fragmentation and ligation of the Illumina P2 adaptor and the cell barcode are performed simultaneously utilizing the Tn5 transposase. To retain only 5′ ends for sequencing, the 3′ ends are digested by the PvuI restriction enzyme.

(C) Smart-seq2. Polyadenylated RNA is reverse transcribed from an Oligo dT with a PCR primer. The same PCR primer is part of the template-switching oligo added to the 5′ end of the cDNA upon reverse transcription. After PCR amplification, the cDNA is fragmented by tagmentation using the Tn5 transposase. Simultaneously, Tn5 ligates different 5′ and 3′ primers to the fragments. Another round of PCR introduces Nextera-sequencing primers to the ends of the fragments, enabling sequencing with full-length read coverage. However, Smart-seq2 does not allow for the integration of UMIs.

Consortium (ERCC) (Baker et al., 2005) and cover a wide range, from 5% to 40%. However, independent methods such as imaging-based molecule counting in single cells by single-molecule fluorescent in situ hybridization (smFISH) (Raj et al., 2008) yield deviating estimates (Grün et al., 2014). Moreover, the absolute number of transcripts per cell is comparable when sequencing cells of the same type with different methods. The ERCC spike-in RNAs are relatively short in comparison to mammalian genes, have short poly(A) tails, and lack a 5′ cap. It is unclear how much these differences between external and cellular RNA—as well as the fact that the external RNA is not spiked directly into the cell—affect the relative sequencing efficiencies of cellular and spike-in RNA.

Data Analysis of Single-Cell Transcriptome Data
Preprocessing and Read Mapping

In order to retrieve the maximum information from single-cell mRNA sequencing data, a careful experimental design is required (see Box 1). Following sequencing, a number of data processing and filtering steps are recommended to reduce the impact of technical noise. The first analysis step is usually a quality filtering or trimming of the sequencing reads prior to mapping

the reads to a reference database. Standard tools, e.g., fastqc, permit a quality analysis of the sequenced library, and standard mapping tools, such as bwa (Li and Durbin, 2010), allow trimming of low-quality bases from the end of the reads. However, a minimum remaining read length (> 35 bp for mouse or human) after trimming should be required in order to avoid false-positive hits. For the mapping, available standard tools developed for bulk RNA-seq analysis can be used (Garber et al., 2011). However, sequenced cell barcodes, UMIs, and other primer-derived sequences have to be removed from the remaining read to be mapped to the reference database. Usually, one read of a pair contains all of the index information, while the other one can be mapped to the gene models (see Figure 1). In general, reads can be mapped to the genome followed by expression quantification via intersecting the read coverage of the genome with gene model annotations. However, this can lead to a larger number of reads mapping to multiple loci, for instance, due to the existence of inactive pseudogenes. Using the transcriptome as a reference reduces the sequence space and increases the fraction of unique reads. Since non-unique reads can introduce spurious correlations between different genes across single cells, it is advisable to discard these reads prior to analysis.

Box 1. Design of Single-Cell Sequencing Experiments

The power of single-cell sequencing crucially depends on two parameters: the number of cells and the sequencing complexity. These parameters can be controlled by the experimental design and should be chosen according to the goal of the study. The size of the dataset, i.e., the number of cells is important for profiling the cell composition of a sample with high sensitivity. Typically, several hundreds of cells have to be sequenced in order to capture not only abundant, but also rare, cell types. Possible biases that might occur during purification of the single-cell sample due to cell size or other factors have to be considered. Moreover, one should incorporate an estimate for the success rate, since a number of single-cell samples will likely yield only little or no material due to RNA degradation or low amplification efficiency. This estimate can be derived from trial experiments. The second parameter is the library complexity. Since the efficiency of single-cell mRNA sequencing is still limited, it is important to sequence each single cell with sufficient sequencing depth. If transcripts are counted with UMIs, the sequencing depth should be adjusted such that every transcript is sequenced at least three to four times. This ensures that even lowly expressed genes can be quantified and do not drop out due to sampling noise. To determine how many cells can be sequenced at once, e.g., on a single lane of an Illumina sequencing machine, the fraction of reads that can be mapped to the transcriptome has to be taken into account. This fraction is typically lower than 50%, since in most protocols additional abundant contributions can originate from sequencing products containing only primer or adaptor sequences (Grün et al., 2014). For example, assuming that ~10,000 transcripts per cell have been amplified and 50% of the reads can be mapped to the transcriptome, about 2,500 cells can be sequenced on a single lane of an Illumina NextSeq machine with 200 million reads. A fraction of those, typically around 10% to 20%, will not pass the quality filtering. Microfluidic devices like the Fluidigm C1 further provide an image of each cell being processed and allow filtering of wells containing no or more than a single cell.

To avoid batch effects, one should follow general guidelines applicable for bulk sequencing. For instance, single-cell libraries corresponding to different conditions should not be sequenced on separate lanes but, rather, distributed in equal fractions across the same set of lanes.

Due to the low read coverage of the gene body in single-cell sequencing experiments, isoform quantification with standard methods such as Cufflinks (Trapnell et al., 2010) can be problematic. If isoform information is not essential for the study, an ideal strategy is to merge all isoforms of a given gene into a so-called gene locus and quantify the expression of these gene loci. Independent of the reference, it is important to consider specific aspects of the experimental strategy. If sequencing protocols are used that enrich for the 5′ or 3′ end of an mRNA, the quality of the gene annotation can have a huge impact on the sensitivity. Gene models tend to be less reliable at both ends of a transcript, and an experimental strategy for improving 5′ or 3′ end annotation might be beneficial, in particular, for non-standard model organisms. For example, Junker et al. applied an amended CEL-seq protocol to sequence longer reads at low depths on bulk material in order to accurately detect 3′ polyadenylation sites for zebrafish embryos (Junker et al., 2014). Finally, the reference database has to be augmented by sequences representing any spike-in RNA added to the samples.

Expression Quantification and Filtering

In order to arrive at expression levels for all genes, PCR duplicates should be removed. Next, the cell of origin is determined based on the sequenced cell barcode (Figure 2A). If the base-calling quality is not sufficiently high at the cell barcode position within the read, an error-tolerant assignment scheme can be applied by aggregating all barcodes up to a single mismatch away from the perfect sequence. In order to apply this scheme, however, each pair of cell barcodes has to differ in at least two positions. If UMIs are available, the number of different UMIs per gene in each cell has to be converted into a transcript count estimate (Figure 2A) by applying a statistical correction to account for sampling effects (Grün et al., 2014; Kivioja et al., 2012).

Once read or transcript counts have been determined for all cells, it is recommended to filter out cells of low yield (Figure 2B). These samples can arise already prior to or during isolation of the cells, e.g., due to stress or apoptosis, or can occur due to incomplete lysis, RNA degradation, or low

sequencing efficiency of a particular cell. The total number of reads or UMI-derived transcript counts per cell is a first proxy for the sample quality. Applying a threshold to discard cells in the lower tail of the distribution of read or transcript counts, respectively, safeguards against artifacts arising from low-quality cells. The expression of spike-in RNA can be utilized to identify and discard samples of low sequencing efficiency. Since the number of spike-in RNA should be identical for all samples, the identification of low yield samples is straightforward (Figure 2B). On the other hand, a relatively large ratio between transcript or read counts, respectively, of spike-in and cellular RNA reveals cells that contribute little material, e.g., due to RNA degradation or incomplete cell lysis (Figure 2B). The described strategies are only guidelines for filtering, and the exact method strongly depends on the dataset under examination. For example, if the cell volume varies substantially within a dataset, the total transcript count should only be subject to mild filtering, while the transcript count of the spike-in RNA is still a good proxy for the sequencing efficiency and can be used to discard low yield samples.

Data Normalization

For subsequent analysis, an appropriate normalization of the expression data is necessary. In the case of read-based quantification, normalization to transcripts per one million reads (TPM)—if reads are only generated from one end of the transcript—or transcript per one million reads per kilobase of transcript (RPKM)—if reads cover the entire transcript—is appropriate. Alternatively, standard quantification methods like Cufflinks (Trapnell et al., 2010) yield normalized expression values. More refined normalization schemes have been developed for bulk RNA-seq data (Anders and Huber, 2010). Here, derivation of a size factor for each replicate accounts for variability in sequencing depth between replicates, and a similar method can be applied to normalize single-cell data (Brennecke et al., 2013). If transcripts are counted with UMIs, cell-to-cell differences in transcript numbers are to a certain extent biologically meaningful and indicative of variations in the RNA content of a
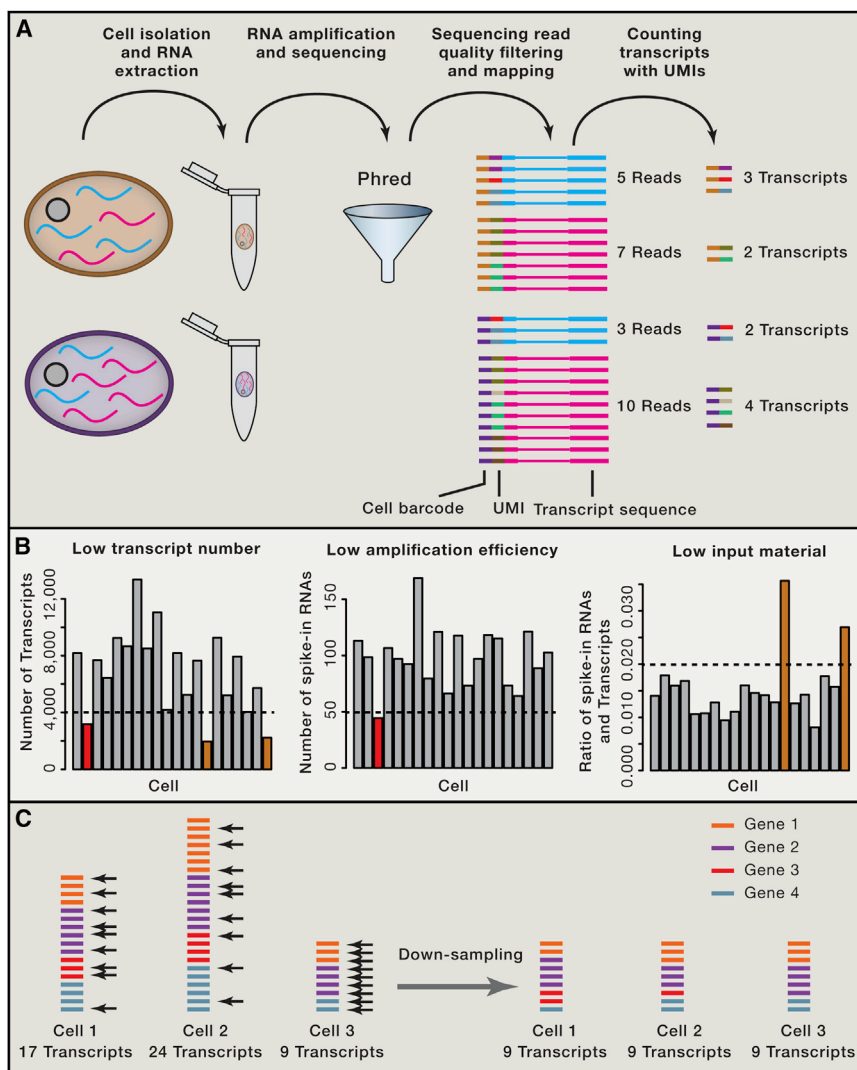
(A) For single-cell sequencing, RNA is isolated from individual cells and, after labeling with cellular barcodes, amplified by PCR or in vitro transcription. Sequencing reads are subject to quality filtering and trimming before mapping to reference sequences representing all genes of the organism. In (A), only two cells with two different genes are shown for simplicity. Amplification bias can distort the relative expression of the two genes and can be eliminated by counting the number of UMIs per genes instead of sequencing reads.

(B) Cells with low yield due to RNA degradation or low sequencing efficiency should be discarded. These cells can be identified based on low total transcript counts (left), which can be explained by low-amplification efficiency (red bar, middle) or low-input material (orange bar, right). The middle panel depicts the total number of spike-in RNA, which should theoretically be the same in all cells. Variations are due to variability in sequencing efficiency. The right panel shows the ratio between spike-in RNA and transcripts of cellular genes. High ratios correspond to reduced amounts of cellular RNA.

(C) Data normalization by down-sampling. The same number of transcripts is randomly picked from each cell. Shown is a toy example with three cells and four different genes.

sampled to a number lower or equal than their actual transcript count. However, for most applications, this approach is preferable since technical artifacts such as batch effects are efficiently eliminated.

Biological Insights from Single-Cell Transcriptome Data
Identification of Cell Types

Perhaps the most important application of single-cell mRNA sequencing is the identification of cell types in a complex mixture. The transcriptome of a cell can be interpreted as a fingerprint revealing its identity. An unbiased screening of randomly sampled cells from a mixture, such as an organ, could therefore reveal the cellular composition of this sample. A number of studies could recover known cell types and identify novel marker genes in diverse systems, for example in the spleen (Jaitin et al., 2014), the lung epithelium (Treutlein et al., 2014), or the retina (Macosko et al., 2015). Another recent landmark paper revealed the complex cellular composition of the mouse hippocampus and uncovered novel cell types (Zeisel et al., 2015). Although these studies convincingly demonstrated that single-cell mRNA sequencing is a powerful method for cell type identification, computational methods to leverage the full complexity within single-cell transcriptome data are just beginning to emerge. Distinguishing cell types in a mixture corresponds to a typical unsupervised learning problem in which data points, in this case given by single-cell transcriptomes, are grouped into clusters reflecting subsets of data points that are more similar to each other than to the remaining data points (Figure 3A). A commonly applied
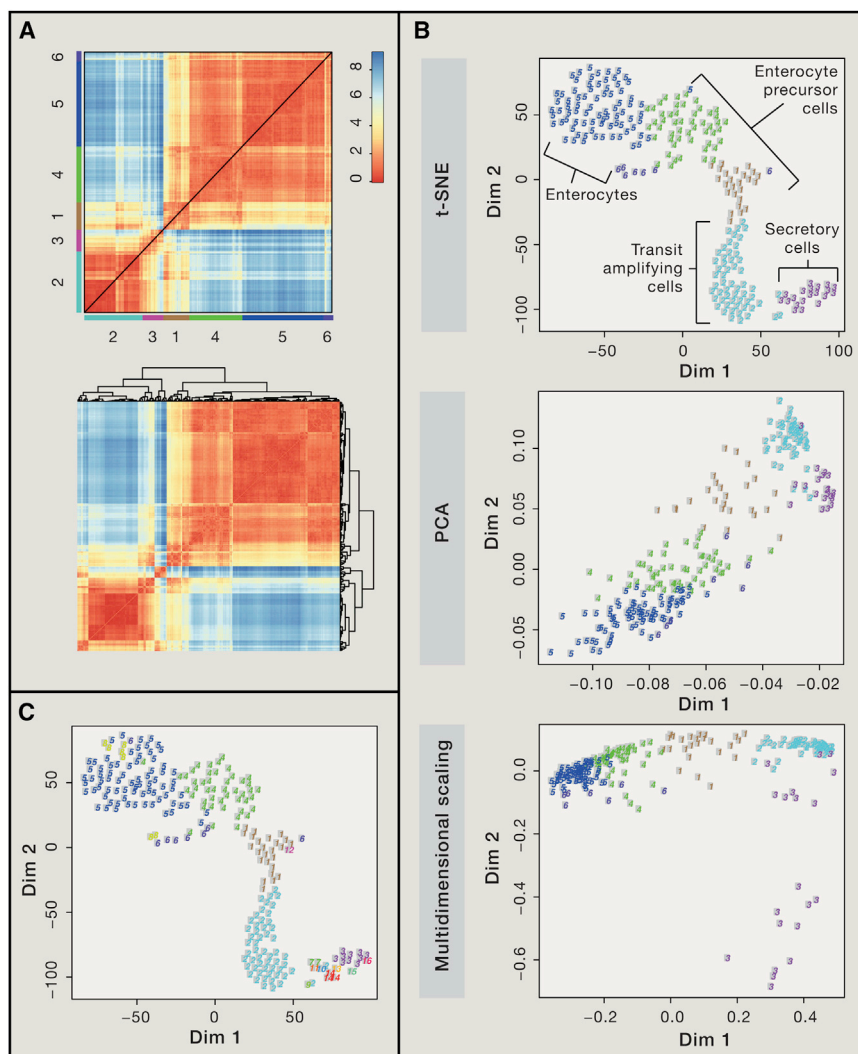
cell. However, cell-to-cell variability in sequencing efficiency and other sources of technical noise contribute to the observed variability. In principle, the technical cell-to-cell variability could be deconvolved with the help of spike-in RNA. The ratio of the number of sequenced spike-in molecules over the number of spike-in molecules added to the cell extract yields a conversion factor. In theory, dividing the number of sequenced transcripts by this conversion factor yields an estimate of the actual number of transcripts in a cell. However, as already discussed, commonly used ERCC spike-in RNA does not provide a good standard for absolute quantification. For most applications, the relative contribution of each gene to the transcriptome will be the relevant readout, and in these cases, simpler normalization schemes apply. One possibility is the normalization of the total transcript count in each cell to the median of the total transcript count across all cells. Alternatively, subsampling of the same number of transcripts from each cell, termed down-sampling (Figure 2C), is more efficient in eliminating technical variability but comes with a loss in complexity since all cells are down-

Figure 3. Single-Cell Sequencing Allows the Inference of Cell Type Composition

(A) Unsupervised learning can be used to distinguish different cell types in a random sample of sequenced cells from a complex mixture. K-means clustering with a cluster number estimated by gap statistics (top) or hierarchical clustering (bottom) based on transcriptome similarity can be used to identify different abundant cell types. All data shown in the figure are derived from 238 random cells isolated from mouse intestinal organoids (Grün et al., 2015).

(B) Dimensional reduction algorithms can be applied for data visualization. The t-SNE method (top) resolves the local structure of the data but tends to group outliers together by their dissimilarity to bigger clusters. PCA (middle) allows visual inspection of data separation along the main axis of variability but can be inconvenient if a larger number of principal components contribute substantial variability. Classical multidimensional scaling achieves dimensional reduction with well-conserved point-to-point distances. Outliers are well separated, but dense clusters tend to be condensed. K-means clusters (A) are highlighted in all of the maps, and intestinal cell types are indicated in the t-SNE map.

(C) The RaceID algorithm identifies rare cell types within more abundant groups separated by k-means clustering. The algorithm can detect cell types represented by only a single cell in the mixture.

within the retina (Macosko et al., 2015) or the hippocampus (Zeisel et al., 2015). These methods take arbitrary similarity measures as input. The most common choices are the Euclidean distance between vectors with expression values for all genes or a correlation-based distance between these vectors, e.g., 1 − Pearson's correlation coefficient.

To identify cell types more systematically, conventional clustering methods can be applied. For instance, hierarchical clustering was used, alone or in combination with PCA, to explore cellular heterogeneity (Patel et al., 2014; Pollen et al., 2014; Treutlein et al., 2014). On the other hand, more sophisticated algorithms have been specifically adjusted for cell type profiling. Jaitin et al. utilized hierarchical clustering to initialize a probabilistic mixture model for cell type classification (Jaitin et al., 2014). Zeisel et al. developed a clustering method based on sorting points into neighborhoods (SPIN) (Tsafrir et al., 2005). In an iterative procedure, an optimal splitting of the cell-to-cell correlation matrix is determined after ordering the expression matrix by cells and genes using SPIN.

A general problem for cell type classification is the presence of confounding factors due to technical and biological variability. The result of any clustering routine has to be carefully examined for batch effects leading to unwanted clustering by experimental batch, sequencing library, or other technical factors. Batch effects can be reduced by normalization strategies such as down-sampling that eliminate differences in complexities between libraries.

visual method is principal component analysis (PCA), which converts a set of correlated variables into a set of orthogonal uncorrelated variables, termed principal components. These principal components are ordered by the fraction of the total variance they explain, and usually only the first two or three principal components are analyzed. Visual inspection of a scatterplot showing the first two principal components can already reveal the main subgroups in the data, i.e., the abundant cell types (Pollen et al., 2014; Shalek et al., 2014; Treutlein et al., 2014). Moreover, a number of algorithms for dimensional reduction exist that can be used to obtain an approximate visualization of the data in two dimensions (Figure 3B). These algorithms take a matrix with all pairwise distances of data points as input and project these points onto a low-dimensional space, trying to preserve the original pairwise distances as much as possible. For example, classical multidimensional scaling was used to visualize intratumoral heterogeneity in glioblastoma (Patel et al., 2014), and t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten and Hinton, 2008) beautifully visualized heterogeneity

However, additional confounding factors can arise from biological heterogeneity such as cell-to-cell differences in the cell-cycle phase. If only cells of a similar size are analyzed, cell sorting can be used to purify cells within a given cell-cycle phase. Otherwise, computational approaches can be used to deconvolve cell-cycle-related variability. A recently published approach utilizes latent variable models to account for the cell cycle and other hidden factors (Buettner et al., 2015). On the other hand, normalization schemes that eliminate absolute cell-to-cell differences in transcript count are often sufficient.

A major challenge for any cell type inference method is the identification of rare cell types. With a frequency of ~1% or less in a sample of sequenced cells from a complex mixture, these cell types typically occur as outliers. Although unsupervised learning methods for outlier identification exist, these approaches oftentimes cannot capture the full complexity of the data. For instance, classifying a variety of different rare cell types in an organ cannot be achieved by these methods (Grün et al., 2015). In a recent study, an algorithm for rare cell type identification (RaceID) was introduced (Grün et al., 2015) that first infers abundant cell types by k-means clustering followed by a systematic outlier screening (Figure 3C). In this step, the cell-to-cell variability of every gene is compared to a background model that accounts for technical and biological noise within a cluster. Cells exhibiting transcript counts with a low p value according to this background model are identified as outliers and are used as new cluster seeds. RaceID was shown to identify rare mouse intestinal cell types with high sensitivity and specificity and discovered novel rare subtypes of the enteroendocrine lineage.

Identification of Marker Genes

Once cell types can be delineated, the data can be mined for specific marker genes to better characterize a cell type and, with the help of cell surface markers or fluorescent reporter genes, allow the purification of a cell type. The discovery of a marker gene requires the identification of differentially expressed genes between the cell type of interest and the remaining cells. For this task, available methods for modeling over-dispersed count statistics in bulk sequencing data, such as DESeq (Anders and Huber, 2010), can be applied. Another probabilistic method, which was developed specifically for single-cell sequencing data, accounts for the relatively high rate of dropout events in these data, i.e., transcripts that escaped reverse transcription and therefore could not be sequenced (Kharchenko et al., 2014).

Inference of Differentiation Dynamics

Related to the cell type inference is the application of single-cell transcriptomics to reveal differentiation pathways. A comparison of single-embryo transcriptomes collected at sub-sequent stages of nematode embryonic development has already revealed insights into gene expression changes underlying the emergence of the three germ layers (Hashimshony et al., 2015). More generally, if a sample is analyzed that contains all differentiation stages of a given cell lineage, a pseudo-temporal ordering of single-cell transcriptomes can be inferred. For example, such a sample can be composed of cells collected at different time points during in vitro differentiation or can be a random sample of a mitotic adult stem cell differentiation system

such as the intestinal epithelium. The general idea is that differentiation is accompanied by continuous temporal changes in gene expression and that ordering of single-cell transcriptomes by similarity reflects the succession of these changes, yielding a pseudo-temporal ordering of single-cell transcriptomes. One existing method termed Monocle combines dimensional reduction with the construction of a minimum spanning tree (Trapnell et al., 2014). Monocle is an unsupervised approach that can infer branching into multiple lineages and was used to elucidate gene expression dynamics during differentiation of primary human fibroblasts. Another more recent method relies on the use of diffusion maps to define differentiation trajectories, incorporating the idea that the movement of a cell within the transcriptional landscape follows diffusion-like dynamics (Haghverdi et al., 2015).

Finally, by defining links between gene pairs, e.g., based on the significance of correlation, a variety of network analysis methods can be applied (Ocone et al., 2015).

There is certainly room for further development of computational methods to infer cell lineages. This inference is particularly challenging if the lineage tree segregates into multiple branches, since technical and biological gene expression noise can confound the assignment of a cell to a particular lineage. The single-cell perspective will yield exciting new insights into the impact of gene expression noise on lineage commitment and on the regulation of gene expression noise during differentiation.

Measuring Gene Expression Noise

Another application of single-cell mRNA sequencing is the investigation of biological gene expression variability, or gene expression noise, in a population of cells. Current models of transcriptional dynamics describe promoter bursting (Figure 4A), where the promoter of a gene switches between an active and an inactive state and, once activated, initiates transcript production at a constant rate (Raj et al., 2006; Raser and O'Shea, 2004). These dynamics imply a variance in transcript levels exceeding the lower limit of pure sampling, i.e., Poissonian noise. Single-cell mRNA sequencing is a suitable method to infer the biological noise and investigate transcriptional parameters on a genome-wide level in a cell population of interest. However, technical noise due to sampling of transcripts to be sequenced from each cell and due to global cell-to-cell variability in sequencing efficiency (Figure 4B) has a substantial contribution to the measured noise levels (Brennecke et al., 2013; Grün et al., 2014). The technical noise component can be quantified, for instance, based on sample-to-sample variability in spike-in RNA levels. After fitting a technical noise model that incorporates sampling noise and global sample-to-sample variability in sequencing efficiency, the technical noise component can be deconvolved from the total noise in order to infer the biological noise component (Figure 4C) (Grün et al., 2014). This approach has been shown to yield precise noise estimates consistent with single-molecule FISH, a highly sensitive imaging-based method for transcript counting (Raj et al., 2008), and can be used, for instance, to measure changes in biological noise between different conditions. Furthermore, given a model of transcriptional bursting, kinetics model parameters such as burst size and burst frequency can be derived from the biological noise estimates (Grün et al., 2014).
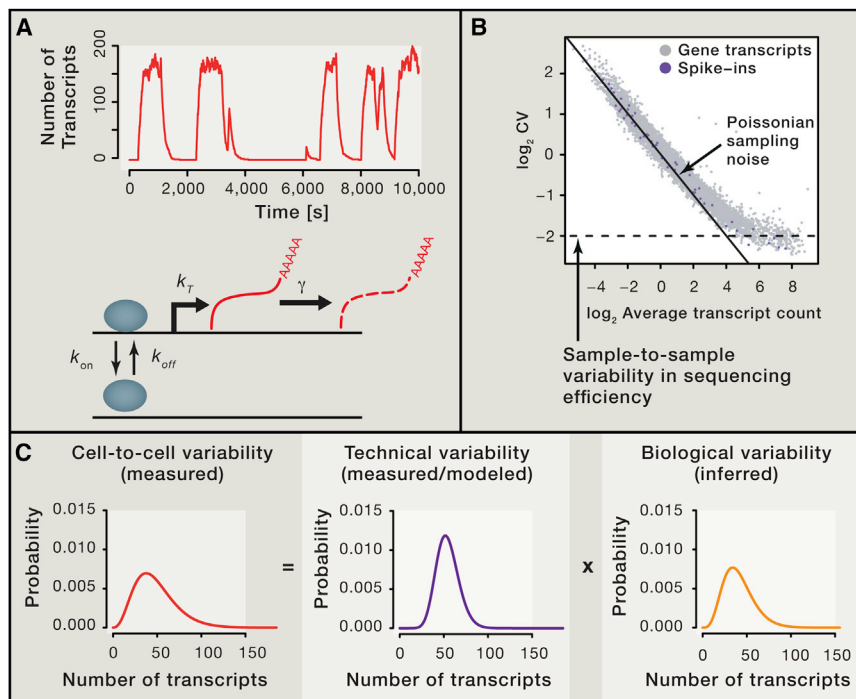
Figure 4. Single-Cell Sequencing Reveals Biological Gene Expression Noise

(A) Transcription is not a time-continuous process. Switching of a gene promoter between an active and an inactive state leads to transcriptional bursting. The kinetic parameters can be estimated from burst size and burst frequency, which can be derived from biological noise estimates measured with single-cell sequencing.

(B) The CV as a function of the mean expression for spike-in RNA or fixed aliquots of cellular RNA reveals sources of technical noise. While sampling noise dominates at low expression, global variability of sequencing efficiency is the major contribution for highly expressed genes.

(C) Technical noise can be modeled and deconvolved from the transcript count distribution measured in cells, yielding good estimates of the actual biological noise (Grün et al., 2014).

Another method allows the identification of highly variable genes by assigning a p value to each gene reflecting to what extent the biological noise exceeds technical variability (Brennecke et al., 2013). This method also relies on technical noise estimates derived from external spike-in RNA.

Investigating Allelic Expression

Single-cell sequencing offers the possibility to study allelic expression on a genome-wide level. If the two alleles of a gene differ by a sufficient number of single-nucleotide polymorphisms (SNPs), transcripts derived from the two alleles can be distinguished by single-cell mRNA sequencing. However, this analysis is highly sensitive to technical noise, i.e., spurious differences in allele frequencies due to sampling effects and stringent controls are required to infer actual biological differences. By analyzing mouse embryos of mixed genetic background, this approach has revealed the presence of abundant random monoallelic expression during preimplantation development and has demonstrated de novo inactivation of the paternal X chromosome (Deng et al., 2014).

Concluding Remarks

The power of single-cell sequencing as a method to characterize the state of a cell across multiple molecular layers has been demonstrated by a number of beautiful studies published during the last few years. Most of the previous research was focused on the investigation of single-cell genomes and transcriptomes. While experimental protocols have improved rapidly, sophisticated computational methods are just beginning to emerge, and in this Primer, we have summarized a number of state-of-the-art methods along with general guidelines covering all analysis stages. We hope that this overview will enable a growing number of researchers to leverage the maximum out of their single-cell sequencing data. The field of single-cell sequencing will keep developing rapidly in the near future and will reveal exciting insights into the regulatory mechanisms that determine the identity of a cell.

REFERENCES

Alexandrov, L.B., and Stratton, M.R. (2014). Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. Curr. Opin. Genet. Dev. *24*, 52–60.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Baker, S.C., Bauer, S.R., Beyer, R.P., Brenton, J.D., Bromley, B., Burrill, J., Causton, H., Conley, M.P., Elespuru, R., Fero, M., et al.; External RNA Controls Consortium (2005). The External RNA Controls Consortium: a progress report. Nat. Methods *2*, 731–734.

Baslan, T., Kendall, J., Rodgers, L., Cox, H., Riggs, M., Stepansky, A., Troge, J., Ravi, K., Esposito, D., Lakshmi, B., et al. (2012). Genome-wide copy number analysis of single cells. Nat. Protoc. *7*, 1024–1041.

Biesecker, L.G., and Spinner, N.B. (2013). A genomic view of mosaicism and human disease. Nat. Rev. Genet. *14*, 307–320.

Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., and Heisler, M.G. (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat. Methods *10*, 1093–1095.

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat. Biotechnol. *33*, 155–160.

Chattopadhyay, P.K., and Roederer, M. (2012). Cytometry: today's technology and tomorrow's horizons. Methods *57*, 251–258.

Cheung, V.G., and Nelson, S.F. (1996). Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. Proc. Natl. Acad. Sci. USA 93, 14676–14679.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2015). Ensembl 2015. Nucleic Acids Res. 43, D662–D669.

Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., et al. (2002). Comprehensive human genome amplification using multiple displacement amplification. Proc. Natl. Acad. Sci. USA 99, 5261–5266.

Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science 343, 193–196.

Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., and Coleman, P. (1992). Analysis of gene expression in single live neurons. Proc. Natl. Acad. Sci. USA 89, 3010–3014.

Eldar, A., and Elowitz, M.B. (2010). Functional roles for noise in genetic circuits. Nature 467, 167–173.

Falconer, E., Hills, M., Naumann, U., Poon, S.S.S., Chavez, E.A., Sanders, A.D., Zhao, Y., Hirst, M., and Lansdorp, P.M. (2012). DNA template strand sequencing of single-cells maps genomic rearrangements at high resolution. Nat. Methods 9, 1107–1112.

Fan, H.C., Wang, J., Potanina, A., and Quake, S.R. (2011). Whole-genome molecular haplotyping of single cells. Nat. Biotechnol. 29, 51–57.

Garber, M., Grabherr, M.G., Guttman, M., and Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. Nat. Methods 8, 469–477.

Gole, J., Gore, A., Richards, A., Chiu, Y.-J., Fung, H.-L., Bushman, D., Chiang, H.-I., Chun, J., Lo, Y.-H., and Zhang, K. (2013). Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells. Nat. Biotechnol. 31, 1126–1132.

Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. Nat. Methods 11, 637–640.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature 525, 251–255. http://dx.doi.org/10.1038/nature14966.

Haghverdi, L., Buettner, F., and Theis, F.J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics 31, 2989–2998. http://dx.doi.org/10.1093/bioinformatics/btv325.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep. 2, 666–673.

Hashimshony, T., Feder, M., Levin, M., Hall, B.K., and Yanai, I. (2015). Spatio-temporal transcriptomics reveals the evolutionary history of the endoderm germ layer. Nature 519, 219–222.

Iscove, N.N., Barbara, M., Gu, M., Gibson, M., Modi, C., and Winegarden, N. (2002). Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. Nat. Biotechnol. 20, 940–943.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.-B., Lönnerberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. 21, 1160–1167.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods 11, 163–166.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., and Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science 343, 776–779.

Junker, J.P., Noël, E.S., Guryev, V., Peterson, K.A., Shah, G., Huisken, J., McMahon, A.P., Berezikov, E., Bakkers, J., and van Oudenaarden, A. (2014). Genome-wide RNA Tomography in the zebrafish embryo. Cell 159, 662–675.

Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. Nat. Methods 11, 740–742.

Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., and Taipale, J. (2012). Counting absolute numbers of molecules using unique molecular identifiers. Nat. Methods 9, 72–74.

Klein, C.A., Schmidt-Kittler, O., Schardt, J.A., Pantel, K., Speicher, M.R., and Riethmüller, G. (1999). Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. Proc. Natl. Acad. Sci. USA 96, 4494–4499.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161, 1187–1201.

Leung, M.L., Wang, Y., Waters, J., and Navin, N.E. (2015). SNES: single nucleus exome sequencing. Genome Biol. 16, 55.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595.

Macaulay, I.C., and Voet, T. (2014). Single cell genomics: advances and future perspectives. PLoS Genet. 10, e1004126.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161, 1202–1214.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.

Meyer, L.R., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Kuhn, R.M., Wong, M., Sloan, C.A., Rosenbloom, K.R., Roe, G., Rhead, B., et al. (2013). The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res. 41, D64–D69.

Möhlendick, B., Bartenhagen, C., Behrens, B., Honisch, E., Raba, K., Knoefel, W.T., and Stoecklein, N.H. (2013). A robust method to analyze copy number alterations of less than 100 kb in single cells using oligonucleotide array CGH. PLoS ONE 8, e67031.

Munsky, B., Neuert, G., and van Oudenaarden, A. (2012). Using gene expression noise to understand gene regulation. Science 336, 183–187.

Nagaoka, S.I., Hassold, T.J., and Hunt, P.A. (2012). Human aneuploidy: mechanisms and new insights into an age-old problem. Nat. Rev. Genet. 13, 493–504.

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D., et al. (2011). Tumour evolution inferred by single-cell sequencing. Nature 472, 90–94.

Nielsen, R., Paul, J.S., Albrechtsen, A., and Song, Y.S. (2011). Genotype and SNP calling from next-generation sequencing data. Nat. Rev. Genet. 12, 443–451.

Ocone, A., Haghverdi, L., Mueller, N.S., and Theis, F.J. (2015). Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. Bioinformatics 31, i89–i96.

Ottolini, C.S., Newnham, L.J., Capalbo, A., Natesan, S.A., Joshi, H.A., Cimadomo, D., Griffin, D.K., Sage, K., Summers, M.C., Thornhill, A.R., et al. (2015). Genome-wide maps of recombination and chromosome segregation in human oocytes and embryos show selection for maternal recombination rates. Nat. Genet. 47, 727–735.

Pan, X., Durrett, R.E., Zhu, H., Tanaka, Y., Li, Y., Zi, X., Marjani, S.L., Euskirchen, G., Ma, C., Lamotte, R.H., et al. (2013). Two methods for full-length RNA sequencing for low quantities of cells and single cells. Proc. Natl. Acad. Sci. USA 110, 594–599.

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 344, 1396–1401.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Pollen, A.A., Nowakowski, T.J., Shuga, J., Wang, X., Leyrat, A.A., Lui, J.H., Li, N., Szpankowski, L., Fowler, B., Chen, P., et al. (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nat. Biotechnol. *32*, 1053–1058.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, e309.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–879.

Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. *30*, 777–782.

Raser, J.M., and O'Shea, E.K. (2004). Control of stochasticity in eukaryotic gene expression. Science *304*, 1811–1814.

Saliba, A.-E., Westermann, A.J., Gorski, S.A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res. *42*, 8845–8860.

Sasagawa, Y., Nikaido, I., Hayashi, T., Danno, H., Uno, K.D., Imai, T., and Ueda, H.R. (2013). Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. Genome Biol. *14*, R31.

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236–240.

Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363–369.

Snijder, B., and Pelkmans, L. (2011). Origins of regulated cell-to-cell variability. Nat. Rev. Mol. Cell Biol. *12*, 119–125.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382.

Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K., and Surani, M.A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. Cell Stem Cell *6*, 468–478.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature *509*, 371–375.

Tsafrir, D., Tsafrir, I., Ein-Dor, L., Zuk, O., Notterman, D.A., and Domany, E. (2005). Sorting points into neighborhoods (SPIN): data analysis and visualization by ordering distance matrices. Bioinformatics *21*, 2301–2308.

Van der Aa, N., Cheng, J., Mateiu, L., Zamani Esteki, M., Kumar, P., Dimitriadou, E., Vanneste, E., Moreau, Y., Vermeesch, J.R., and Voet, T. (2013). Genome-wide copy number profiling of single cells in S-phase reveals DNA-replication domains. Nucleic Acids Res. *41*, e66.

Van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. *9*, 2570–2605.

Venkatraman, E.S., and Olshen, A.B. (2007). A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics *23*, 657–663.

Voet, T., Kumar, P., Van Loo, P., Cooke, S.L., Marshall, J., Lin, M.-L., Zamani Esteki, M., Van der Aa, N., Mateiu, L., McBride, D.J., et al. (2013). Single-cell paired-end genome sequencing reveals structural variation per cell cycle. Nucleic Acids Res. *41*, 6119–6138.

Wang, J., Fan, H.C., Behr, B., and Quake, S.R. (2012). Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. Cell *150*, 402–412.

Wang, Y., Waters, J., Leung, M.L., Unruh, A., Roh, W., Shi, X., Chen, K., Scheet, P., Vattathil, S., Liang, H., et al. (2014). Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature *512*, 155–160.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.

Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., and Arnheim, N. (1992). Whole genome amplification from a single cell: implications for genetic analysis. Proc. Natl. Acad. Sci. USA *89*, 5847–5851.

Zhang, C., Zhang, C., Chen, S., Yin, X., Pan, X., Lin, G., Tan, Y., Tan, K., Xu, Z., Hu, P., et al. (2013). A single cell level based method for copy number variation analysis by low coverage massively parallel sequencing. PLoS ONE *8*, e54236.

Zhao, M., Wang, Q., Wang, Q., Jia, P., and Zhao, Z. (2013). Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics *14* (*Suppl 11*), S1.

Zong, C., Lu, S., Chapman, A.R., and Xie, X.S. (2012). Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science *338*, 1622–1626.

Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer

Fa-Xing Yu,[1,*] Bin Zhao,[2] and Kun-Liang Guan[3,*]
[1]Children's Hospital and Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China
[2]Life Sciences Institute and Innovation Center for Cell Signaling Network, Zhejiang University, Hangzhou, Zhejiang 310058, China
[3]Department of Pharmacology and Moores Cancer Center, University of California, San Diego, La Jolla, CA 92093, USA
*Correspondence: fxyu@fudan.edu.cn (F.-X.Y.), kuguan@ucsd.edu (K.-L.G.)
http://dx.doi.org/10.1016/j.cell.2015.10.044

Two decades of studies in multiple model organisms have established the Hippo pathway as a key regulator of organ size and tissue homeostasis. By inhibiting YAP and TAZ transcription co-activators, the Hippo pathway regulates cell proliferation, apoptosis, and stemness in response to a wide range of extracellular and intracellular signals, including cell-cell contact, cell polarity, mechanical cues, ligands of G-protein-coupled receptors, and cellular energy status. Dysregulation of the Hippo pathway exerts a significant impact on cancer development. Further investigation of the functions and regulatory mechanisms of this pathway will help uncovering the mystery of organ size control and identify new targets for cancer treatment.

The emergence of multicellular organisms is an evolutionary milestone. Among the most fundamental mechanisms supporting multicellularity are those that ensure the proper size and shape of tissues and organs to meet the need of functionality. However, despite intensive investigations into the underlying principles behind a "preset" size of organs, we are far from having a clear picture of this basic question in developmental biology. Nevertheless, investigations of the Hippo pathway on organ size control have shed light into this mystery (Halder and Johnson, 2011; Pan, 2010; Yu and Guan, 2013).

In 1995, two studies in *Drosophila* discovered that deletion of *Warts* (*wts*) gene resulted in dramatic overgrowth of multiple tissues (Justice et al., 1995; Xu et al., 1995). Several years later, a flurry of studies showed that *Salvador* (*sav*) (Kango-Singh et al., 2002; Tapon et al., 2002), *Hippo* (*hpo*) (Harvey et al., 2003; Jia et al., 2003; Pantalacci et al., 2003; Udan et al., 2003; Wu et al., 2003), and *Mob as tumor suppressor* (*mats*) (Lai et al., 2005) mutant *Drosophila* phenocopy *wts* mutants with regards to tissue overgrowth. Hpo, Sav, Wts, and Mats interact genetically and physically, and the remarkable organ size phenotype elicited by their mutation is unprecedented in other established developmental pathways; thus, they were grouped into a new signaling module—the Hippo pathway—named after the enormous size of *hpo* mutant organs, which resembles that of a hippopotamus. *Yorkie* (*yki*), the key functional effector of the Hippo pathway in organ size regulation, was soon discovered in a screen for Wts-interacting proteins (Huang et al., 2005).

The Hippo pathway is highly conserved in mammals. The mammalian orthologs of *Hpo*, *Sav*, *Wts*, *Mats*, and *Yki* are *Mammalian sterile 20-like 1/2* (*MST1/2*, also called *STK4/3*), *Salvador* (*SAV1*), *Large tumor suppressor homolog 1/2* (*LATS1/2*), *MOB kinase activator 1A/B* (*MOB1a/b*), and *Yes-associated protein* (*YAP*)/*transcriptional co-activator with PDZ binding motif* (*TAZ,* also called *WWTR1*), respectively (Halder and Johnson, 2011; Pan, 2010; Yu and Guan, 2013) (Table 1). The Hippo

pathway quickly attracted broad attention due to its remarkable potency in regulating organ size, as well as its apparent relevance to tissue regeneration and cancer. Here, with an emphasis on recent developments, we review the current understanding of the organization, regulation, and function of the Hippo pathway and discuss some key open questions.

YAP/TAZ and Yki as Major Effectors of Hippo Signaling

In *Drosophila*, deleting *yki* suppresses the overgrowth phenotypes of *hpo*, *sav*, or *wts* mutants (Huang et al., 2005). In mice, deleting *Yap* also diminishes the overgrowth phenotypes caused by deficiency of *Mst1/2* or other upstream regulators (Zhang et al., 2010; Zhou et al., 2011). Thus, Yki and YAP/TAZ are the evolutionarily conserved key effectors of the Hippo pathway.

Yki and YAP/TAZ are believed to mediate the biological functions of the Hippo pathway by regulating gene transcription. As transcriptional co-activators, Yki and YAP/TAZ cannot bind DNA directly, and they must interact with DNA-binding transcription factors to regulate target gene expression. In *Drosophila*, Scalloped (Sd) is a transcription factor and key partner of Yki that mediates the function of Yki in tissue growth (Goulev et al., 2008; Wu et al., 2008; Zhang et al., 2008; Zhao et al., 2008). Similarly, in mammalian cells, the TEAD family transcription factors (TEAD1–4, orthologs of Sd) are key partners for YAP (Zhao et al., 2008). Several lines of evidence indicate that TEADs are the major partners of YAP in transcriptional regulation. For instance, a TEAD-binding-deficient YAP mutant lost its ability to induce transcription of known YAP target genes (Zhao et al., 2008), and knockin of the TEAD-binding-deficient YAP (Y94A mutant) phenotypically mimics YAP knockout in the skin and heart (Schlegelmilch et al., 2011; von Gise et al., 2012). In addition, the majority of YAP and TEAD occupied sites in the genome are shared (Stein et al., 2015; Zanconato et al., 2015; Zhao et al., 2008), and when TEAD is fused with a VP16

Table 1. Hippo Pathway Components and Major Functions

Drosophila	Mammals	Major Functions in Hippo Pathway
Hippo (Hpo)	MST1/2	Phosphorylates LATS1/2, MOB1, and SAV1, leading to LATS1/2 activation
Salvador (Sav)	SAV1	Interacts with MST1/2, promotes phosphorylation of LATS1/2 by MST1/2
Warts (Wts)	LATS1/2	Phosphorylates and inactivates YAP/TAZ
Mats	MOB1A/B	Scaffold protein of LATS1/2
Yorkie (Yki)	YAP/TAZ	Transcription co-activator, major effector of the Hippo pathway
Scalloped (Sd)	TEAD1-4	Transcription factors that mediate the effect of YAP/TAZ
Tgi	VGLL4	Competes with YAP/TAZ for TEADs, inhibits YAP/TAZ functions
misshapen (Msn)	MAP4K4/6/7	Phosphorylates and activates LATS1/2
Merlin (Mer)	Merlin/NF2	May form a complex and mediates upstream signals (from plasma membrane) to MST1/2; NF2 may bring LATS1/2 to plasma membrane and facilitate its activation by MST1/2
Kibra	KIBRA	
Expanded (Ex)	FRMD6?	
	AMOT	Sequesters YAP/TAZ to cell junctions, binding and indirectly activating LATS1/2; a substrate of LATS1/2

transactivation domain, it produces a gene expression profile similar to that induced by active YAP (Ota and Sasaki, 2008). TEAD1–4 or Sd can bind to a consensus motif similar to the GTIIC sequence (TGGAATGT or ACATTCCA), and transcription reporters under control of GTIIC concatemers are now widely used to measure Hippo pathway activity (Dupont et al., 2011; Mohseni et al., 2014; Ota and Sasaki, 2008). Besides TEAD1–4, YAP/TAZ have also been shown to interact with other transcription factors, including Smad, RUNX1/2, p63/p73, and ErbB4 (reviewed in Varelas, 2014), although the functional significance of these transcription factors in Hippo pathway is less clear.

A strong transcriptional activation domain is present in YAP/TAZ but absent in Yki. Nevertheless, human YAP can rescue the lethality resulting from Hippo pathway hyperactivation in *Drosophila*, indicating a functional conservation (Huang et al., 2005). Genome-wide assessment of chromatin binding status reveals that, in addition to occupancy at proximal promoters of target genes, YAP and TEAD largely exert their transcriptional activity by interacting with distal enhancers, suggesting that YAP, and probably TAZ and Yki, may regulate transcription via multiple mechanisms, including recruitment of general transcription factors, modification of epigenetic markers, and modulation of chromatin structure (Lian et al., 2010; Stein et al., 2015; Zanconato et al., 2015) (Figure 1). Indeed, recent evidence shows that Yki can interact with the Brahma complex, GAGA factors, nuclear receptor coactivator 6 (NCoA6, a subunit of the Trithorax-related histone methyltransferase), and the Mediator complex (Jin et al., 2013; Oh et al., 2013; Qing et al., 2014), and TAZ can interact with the chromatin-remodeling complex SWI/SNF (Skibinski et al., 2014).

The transcriptional activity of Yki and YAP/TAZ is regulated in the nucleus by Tondu-domain-containing growth inhibitor (Tgi) and Vestigial-like family member 4 (VGLL4, an ortholog of Tgi). Tgi can directly compete with Yki for Sd binding, resulting in inhibition of Yki-regulated transcription (Guo et al., 2013; Koontz et al., 2013). When Hippo signaling is on, Tgi and Sd form a complex, leading to transcriptional repression; on the contrary, when Hippo signaling is off, Yki enters the nucleus and displaces Tgi

from Sd, leading to expression of Yki target genes and tissue growth (Koontz et al., 2013). In mammals, VGLL4 similarly competes with YAP/TAZ for TEAD binding (Jiao et al., 2014; Zhang et al., 2014b) (Figure 1). However, whether YAP/TAZ functions simply by relieving a default repression of TEAD by VGLL4 has not been demonstrated. Interestingly, the expression of VGLL4 appears to be repressed by miR-130a—a microRNA (miRNA) that is directly induced by YAP, leading to amplification of YAP activity (Shen et al., 2015). A similar mechanism is also present in *Drosophila*, in which *Bantam*—a well-known Yki-induced miRNA (Nolo et al., 2006; Thompson and Cohen, 2006), can repress expression of Tgi (Shen et al., 2015).

Although their role as transcriptional co-activators is widely appreciated, YAP/TAZ and Yki may also repress expression of certain genes when bound to specific factors. For instance, YAP/TAZ can interact with the nucleosome-remodeling and histone deacetylase (NuRD) complex, resulting in transcriptional repression (Kim et al., 2015). Moreover, YAP has been identified as a regulator for global miRNA biogenesis via modulation of miRNA-processing enzymes Microprocessor or Dicer complex, suggesting a transcription-independent role for YAP (Chaulk et al., 2014; Mori et al., 2014). Hence, YAP/TAZ and Yki may regulate gene expression via multiple mechanisms.

Gene expression signatures under YAP/TAZ and Yki hyperactivation (or ectopic expression) in *Drosophila* and different mammalian cell types have been profiled by independent studies. However, the overlap between these different gene profiling studies is not high, suggesting that YAP/TAZ and Yki may regulate target gene expression in a tissue- or cell-type-specific manner. In *Drosophila*, some common Yki target genes are *Bantam*, *diap1*, and *cyclin E*, which may mediate the effect of Yki in inhibiting cell death and promoting cell proliferation (Huang et al., 2005; Nolo et al., 2006; Tapon et al., 2002; Thompson and Cohen, 2006). In mammals, connective tissue growth factor (CTGF) is a commonly used marker of YAP activation (Zhao et al., 2008), and many genes involved in cell proliferation, cell adhesion, and cell migration have also been identified as YAP targets (Stein et al., 2015). It is likely that a panel of genes work together to exert biological functions of YAP/TAZ.
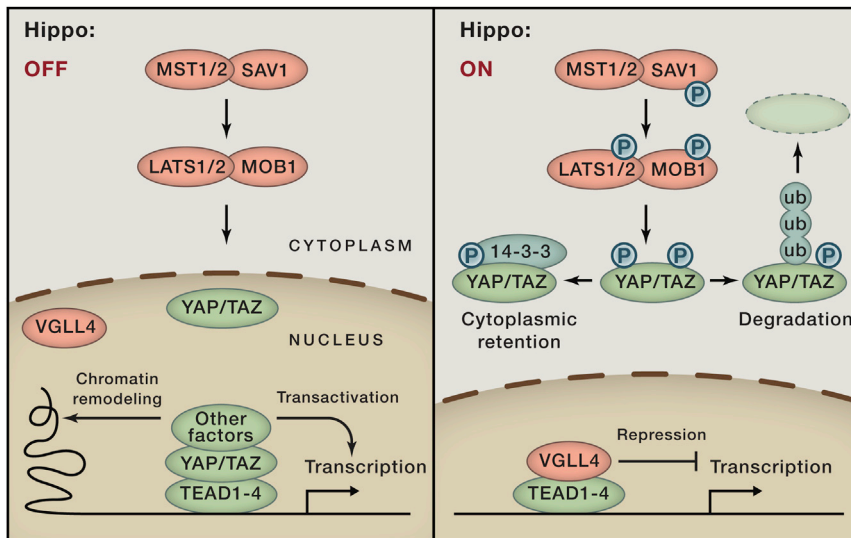
Figure 1. Inhibition of YAP/TAZ Transcriptional Coactivators by LATS1/2

(Left) When Hippo signaling is off, YAP/TAZ enter the nucleus, compete with VGLL4 for TEADs, and recruit other factors to induce gene transcription. YAP/TAZ may bind proximal promoters or distal enhancers of target genes to induce transcription. (Right) When Hippo signaling is on, YAP/TAZ are phosphorylated by LATS1/2 on multiple sites, resulting in interaction with 14-3-3 and cytoplasmic retention; phosphorylation also leads to YAP/TAZ poly-ubiquitination and degradation. VGLL4 interacts with TEADs and represses target gene transcription.

et al., 2002) (Figure 1). In *Drosophila*, Wts is regulated by Hpo, Sav, and Mats by a similar mechanism (Wehr et al., 2013).

Merlin (Mer), Expanded (Ex), and Kibra, three cell-cortex-localized and cytoskeleton-interacting proteins, may function as a scaffold for core Hippo components at the apical domain for activation, as Sav, Hpo, and Wts have been shown to physically interact with Ex/Mer/Kibra (Baumgartner et al., 2010; Genevet et al., 2010; Hamaratoglu et al., 2006; Yu et al., 2010) (Table 1). In addition, the effect of Ex/Mer/Kibra on the Hippo pathway may be mediated by Tao kinase 1 (Tao-1), which can directly phosphorylate and activate Hpo (Boggiano et al., 2011; Poon et al., 2011). In mammalian cells, Neurofibromin 2 (NF2, Mer ortholog) appears to play a more direct role in regulating LATS1/2 activity: it can directly interact with LATS1/2 and recruit LATS1/2 to the plasma membrane for activation by MST1/2 (Yin et al., 2013; Zhang et al., 2010) (Figures 2 and 3).

Hpo or MST1/2 are not absolutely required for regulation of Wts or LATS1/2. It has been observed that, in mouse embryonic fibroblast (MEF) cells, MST1/2 double knockout did not abolish YAP phosphorylation, suggesting the existence of additional Hippo-like activity (Zhou et al., 2009). Indeed, a recent study in *Drosophila* has identified Misshapen (Msn) as another kinase responsible for Wts activation. This mechanism is also conserved in mammals, as MAP4K4 (Msn ortholog) overexpression promotes phosphorylation of LATS1/2 (Li et al., 2014), and MAP4K4 knockdown induces activity of a YAP reporter (Mohseni et al., 2014). In addition to MAP4K4, two recent studies revealed that many MAP4K family kinases, including MAP4K1/2/3/5 (Happyhour in *Drosophila*) and MAP4K4/6/7 (Msn in *Drosophila*), can directly phosphorylate and activate LATS1/2 (Meng et al., 2015; Zheng et al., 2015). These kinases, together with MST1/2, may regulate LATS1/2 activity in a tissue- and signal-specific manner. It is also possible that additional kinases, especially some STE20 family members, may activate LATS1/2 in response to different upstream signals or in different tissue contexts (Figures 2 and 3).

YAP/TAZ have also been shown to be phosphorylated by many other kinases such as cyclin-dependent kinase 1 (CDK1), Jun N-terminal kinases (JNK), homeodomain-interacting protein kinases (HIPK), ABL, and Src family tyrosine kinases (reviewed in Varelas, 2014), suggesting that YAP/TAZ can be

MST1/2 and LATS1/2 Kinases Restrict YAP Activity

In the Hippo pathway, LATS1/2 directly phosphorylate and inhibit YAP/TAZ. Interestingly, YAP has five LATS1/2 target consensus motifs (HXRXXS), four of which are conserved in TAZ (Zhao et al., 2010). Phosphorylation of YAP on serine 127 (S127) generates a 14-3-3 binding site, and binding with 14-3-3 sequesters YAP in the cytoplasm (Dong et al., 2007; Zhao et al., 2007). In addition, phosphorylation of YAP on serine 381 (S381) triggers a subsequent phosphorylation by casein kinase 1 (CK1δ/ε) and activation of a phosphodegron, resulting in recruitment of SCF^beta-TRCP E3 ligase, ubiquitination, and proteasomal degradation of YAP (Zhao et al., 2010). Thus, through regulating both YAP subcellular localization and protein stability, LATS1/2 ensures a spatial and temporal control of YAP activity (Figure 1). TAZ is regulated by LATS1/2 in a similar fashion, although degradation plays a more prominent role in TAZ regulation possibly due to an additional phosphodegron at its N terminus (Lei et al., 2008; Liu et al., 2010a). The subcellular localization of Yki is regulated similarly by Wts phosphorylation and 14-3-3 binding. However, the phosphodegron and the phosphorylation-mediated degradation mechanisms are not conserved in Yki (Huang et al., 2005; Oh and Irvine, 2008). In addition, YAP and Yki have also been shown to be degraded via the autolysosomal pathway (Kwon et al., 2013; Liang et al., 2014), suggesting a potential role of YAP and Yki in vesicular membrane dynamics and related cellular processes such as autophagy.

LATS1/2 are activated by MST1/2 through several mechanisms. MST1/2 phosphorylate LATS1/2 at the C-terminal hydrophobic motif, which promotes LATS1/2 autophosphorylation at its activation loop. Furthermore, phosphorylation of MOB1 by MST1/2 enhances MOB1 interaction with the autoinhibitory domain of LATS1/2, leading to full activation of LATS1/2 (Callus et al., 2006; Chan et al., 2005; Praskova et al., 2004; Wu et al., 2003). In addition, SAV1 is also phosphorylated by MST1/2, and SAV1 functions as a partner of MST1/2 in promoting LATS1/2 phosphorylation (Callus et al., 2006; Tapon
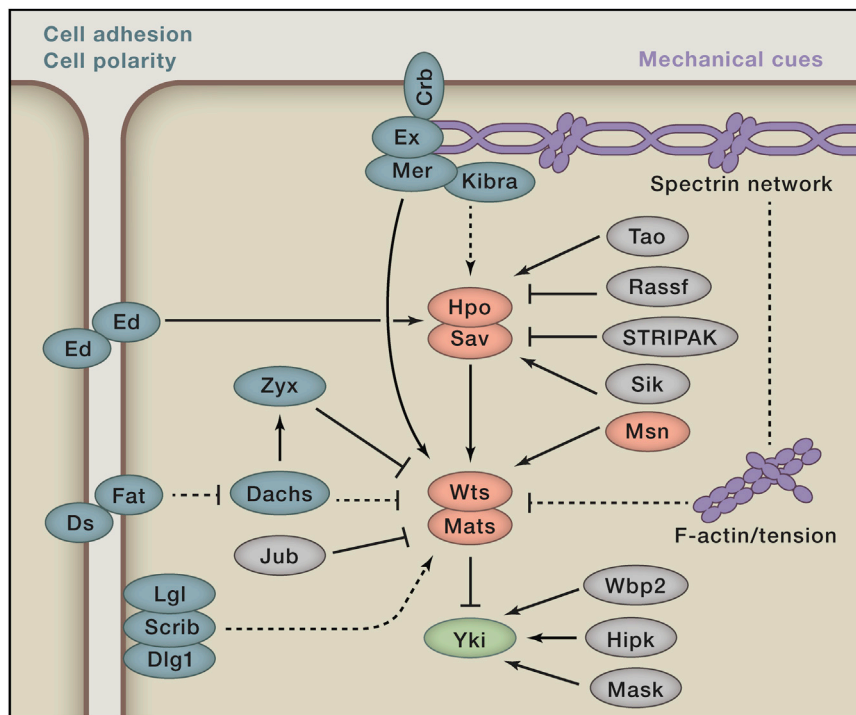
Cell Contact and Mechanical Cues Regulate Hippo Signaling

Cells in solid tissues communicate with neighboring cells and their extracellular matrix (ECM) and perceive constant physical signals from their local environment. Cell-cell contact was discovered as the first signal regulating the Hippo pathway (Zhao et al., 2007). In a sparse culture, YAP/TAZ are primarily localized in the nucleus to promote target gene transcription and cell proliferation. On the contrary, at high cell density, YAP/TAZ are primarily cytoplasmic, corresponding to growth inhibition. It is known that a cell ceases to proliferate following physical contact with surrounding cells, and loss of cell contact inhibition is an indicator of oncogenic transformation. Thus, regulation of YAP/TAZ by cell density suggests a critical role for the Hippo pathway in contact inhibition, tissue growth, and tumorigenesis.

Mechanical cues, such as ECM stiffness and cell geometry, are also potent regulators of YAP/TAZ (Dupont et al., 2011) (Figure 3). When cells are grown on a stiff matrix or are spread across a large surface, YAP/TAZ are activated. In contrast, when cells are seeded on a soft matrix or are compressed into a small area, YAP/TAZ are inactivated. ECM stiffness and cell geometry are important for cell proliferation and differentiation, and YAP/TAZ activity plays a role in these cellular processes. Cell geometry has also been proposed as a mechanism underlying Hippo pathway regulation by cell density: at low density, cells are flat and spread, leading to YAP activation; whereas at high density, cells adopt a round and compact geometry, resulting in YAP inactivation (Aragona et al., 2013; Wada et al., 2011). As further support for a role of YAP/TAZ in mechanosensing, mechanical strain and shear stress have been shown to stimulate YAP/TAZ, and YAP activation is required for mechanical strain-induced cell-cycle reentry (Benham-Pyle et al., 2015; Codelia et al., 2014; Kim et al., 2014). In a medaka fish model, YAP also mediates changes in three-dimensional body shape in response to tissue tension (Porazinski et al., 2015).

The attachment of cells to the ECM is an important mechanism for maintaining cell survival, and cells normally undergo anoikis when detached from the ECM. The attachment status of a cell to its ECM can also regulate Hippo pathway activity. YAP is
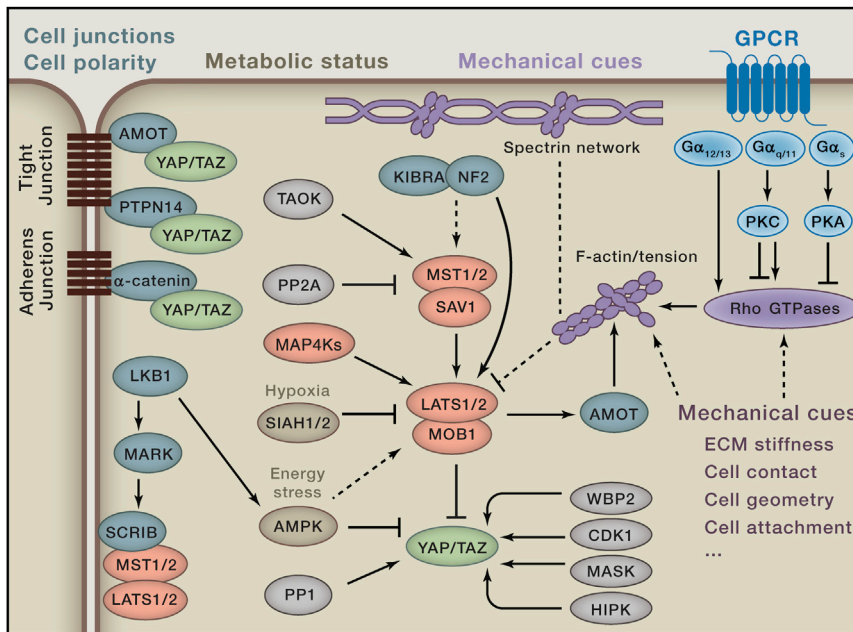
regulated by mechanisms independent of Hippo pathway kinases.

Cell Polarity and Cell Adhesion Regulate Hippo Signaling

In searching for upstream regulators of the Hippo pathway, many proteins involved in cell polarity and cell adhesion have been identified. Echinoid (Ed), a cell adhesion molecule in *Drosophila*, can interact with and stabilize Sav and leads to activation of Hpo (Yue et al., 2012). In mammalian cells, several proteins at adherens and tight junctions, such as Angiomotin (AMOT), protein tyrosine phosphatase non-receptor type 14 (PTPN14), and α-catenin, can sequester YAP/TAZ at cell junctions (reviewed in Yu and Guan, 2013). Therefore, cell adhesion and formation of intercellular junctions serve as a mechanism to repress YAP/TAZ transcriptional activity (Figures 2 and 3).

Crumbs (Crb), a component of apical-basal polarity, interacts with Ex, which is critical for the apical localization of Ex/Mer/Kibra complex (Chen et al., 2010; Ling et al., 2010; Robinson et al., 2010). In addition, Scribble (SCRIB) interacts with both MST1/2 and LATS1/2, thus promoting LATS1/2 activation (Cordenonsi et al., 2011; Mohseni et al., 2014). Fat, a protocadherin that plays a key role in planar cell polarity, also activates the Hippo pathway, possibly through regulating Ex protein levels and localization (Bennett and Harvey, 2006; Silva et al., 2006; Tyler and Baker, 2007; Willecke et al., 2006) (Figure 2). However, mammalian Fat orthologs do not seem to be major regulators of the Hippo pathway (Sharma and McNeill, 2013). It is noteworthy that the link between cell polarity and the Hippo pathway may be indirect, and some proteins, such as Fat, may regulate cell polarity and Hippo signaling via different mechanisms (Matakatsu and Blair, 2012).

Figure 3. Regulation of the Hippo Pathway in Mammals

The Hippo pathway is regulated by diverse signals: (1) determinants of cell polarity and cell-cell junctions, such as SCRIB, which interacts with MST1/2 and LATS1/2, AMOT, PTPN14, and α-catenin, which can sequester YAP/TAZ to cell junctions; (2) mechanical cues, such as stiffness, cell contact, cell geometry, and cell attachment status that regulate the Hippo pathway by modulating activity of Rho GTPases, remodeling the actin cytoskeleton, or altering cellular tension; both apical and basolateral spectrin networks may function as sensors for mechanical cues in Hippo pathway regulation; (3) soluble factors, especially ligands for GPCRs, regulate LATS1/2 likely through Rho GTPases and actin dynamics; (4) metabolic status, such as cellular energy and oxygen stress, also regulate Hippo signaling; many other proteins such as protein phosphatase 2A (PP2A), protein phosphatase 1 (PP1), WBP2, CDK1, MASK, and HIPK can also regulate activities of different Hippo pathway components (refer to Varelas, 2014). Arrows, blunt ends, and dashed lines indicate activation, inhibition, and indirect regulation, respectively.

activated during the process of cell attachment but inactivated when cells are detached (Zhao et al., 2012). Expression of constitutively active YAP promotes survival of detached cells, suggesting that cancer cells with high YAP activity may escape anoikis and undergo metastasis. In support of this notion, LATS1/2 expression was found to be selectively reduced in metastatic, but not primary, prostate tumors (Zhao et al., 2012).

Soluble Factors Regulate Hippo Signaling

The primary function of YAP/TAZ is to promote growth, and many mitogenic hormones and growth factors act through G-protein-coupled receptors (GPCR) to induce cell proliferation (Figure 3). Ligands signal through GPCRs coupled to $G\alpha_{12/13}$ or $G\alpha_{q/11}$, such as lysophosphatidic acid, thrombin, angiotensin II, and estrogen, and can activate YAP/TAZ; in contrast, ligands signal through $G\alpha_s$-coupled GPCRs and protein kinase A (PKA), such as epinephrine and glucagon, can repress YAP/TAZ activity (Kim et al., 2013; Miller et al., 2012; Mo et al., 2012; Wennmann et al., 2014; Yu et al., 2013a; Yu et al., 2012; Zhou et al., 2015). Interestingly, activation of protein kinase C (PKC) by $G\alpha_{q/11}$ can either activate or inhibit YAP, with conventional PKC activating YAP and novel PKC inhibiting YAP (Gong et al., 2015). The remarkable differential functions of PKC in YAP regulation provide a mechanism to explain some of the cell-type-specific responses to PKC activation. GPCR is the largest family of membrane receptors mediating diverse physiological or pathological responses. The demonstration of Hippo regulation by GPCRs links the Hippo pathway to a wide range of upstream signals and biological functions.

The Wnt/β-catenin pathway is a key signaling cascade in development and carcinogenesis. A destruction complex including Axin, adenomatous polyposis coli (APC), and glycogen synthase kinase-3 (GSK3) causes constant degradation of β-catenin. Wnt stimulation disrupts the destruction complex and leads to accumulation of β-catenin. Interestingly, recent studies suggest that YAP/TAZ are also activated by diverse Wnt family ligands. YAP/TAZ have been shown to be components of the destruction complex and are regulated by Wnt in a fashion similar to that of β-catenin (Azzolin et al., 2012, 2014). However, a recent study suggests that Wnt activates YAP/TAZ via Frizzled (a GPCR-like Wnt receptor), $G\alpha_{12/13}$, Rho GTPases, and LATS1/2 (Park et al., 2015). In addition, APC has been shown to act as a scaffold protein for SAV1 and LATS1, and *Apc* deletion leads to YAP activation and tumorigenesis (Cai et al., 2015). More studies may be needed to verify the mechanism of YAP/TAZ activation by Wnt.

Epidermal growth factor (EGF) and insulin have also been shown to regulate YAP/Yki activity in cultured mammalian cells and *Drosophila*, and these signals are mediated by the Ras-Raf-MAPK signaling cascade or phosphoinositide-dependent kinase (PDK1) (Fan et al., 2013; Reddy and Irvine, 2013; Straßburger et al., 2012). TAZ is also stabilized upon PI3K activation, which is mediated by direct phosphorylation by GSK3 (Huang et al., 2012). However, no significant effect of EGF and IGF on YAP was observed in several other studies, and YAP activity appeared to be normal in the presence of inhibitors of PI3K or AKT or in PDK1 null embryonic cells (Yu et al., 2012; Zhao et al., 2007). These discrepancies could be due to differences in experimental settings or cell types and should be clarified by future studies.

Effect of Cellular Metabolic Status on Hippo Signaling

Recently, a link between cellular metabolic status and the Hippo pathway has been reported (Figure 3). Under energy deprivation, such as glucose starvation, the AMPK-activated protein kinase (AMPK) can directly phosphorylate YAP at serine 61 (S61) and serine 94 (S94) (Mo et al., 2015; Wang et al., 2015). Phosphorylation of YAP S94 abolishes the interaction between YAP and TEADs, leading to inhibition of YAP activity. In addition, energy stress also inhibits YAP by increasing kinase activity of LATS1/2 either in an AMPK-dependent or -independent

manner (Mo et al., 2015; Wang et al., 2015). AMPK can phosphorylate AMOTL1, which in turn facilitates YAP phosphorylation by LATS1/2 (DeRan et al., 2014). A similar mechanism is also present in *Drosophila*, in which Ampk inactivates Yki and affects cell proliferation in the larval central brain and central nerve cord (Gailite et al., 2015). It appears that, during energy stress, both AMPK and LATS1/2 are unleashed to restrict YAP activity. These findings also suggest that the Hippo pathway may mediate the anticancer effect of metformin, which is known to activate AMPK (DeRan et al., 2014; Mo et al., 2015). Other than AMPK, glucose may also promote YAP/TAZ activity through phosphofructokinase, which stimulates the interaction between YAP/TAZ and TEADs (Enzo et al., 2015).

YAP/TAZ activity has also been linked to oxygen availability. Under hypoxic conditions, the hypoxia-inducible factor 1 (HIF1) stimulates expression of SIAH1/2, two E3 ubiquitin ligases. SIAH1/2 then promote ubiquitination and degradation of LATS2, leading to YAP/TAZ activation (Ma et al., 2015; Xiang et al., 2014) (Figure 3). In addition, HIF1 directly induces transcription of TAZ (Xiang et al., 2014), and YAP interacts with and stabilizes HIF1 to enhance transcription of HIF1 target genes (Ma et al., 2015).

YAP/TAZ are potent stimulators of cell growth and proliferation, which are energy-consuming processes. The regulation of Hippo signaling by AMPK suggests that metabolic status can function as a checkpoint for growth-promoting activity of YAP/TAZ. Under conditions of nutrient deprivation or energy crisis, YAP/TAZ activity needs to be restricted to prevent energy exhaustion caused by anabolic processes. Oxygen also plays a critical role in cellular metabolism, and hypoxia is involved in different pathological processes such as cancer. The link between hypoxia and YAP/TAZ activity indicates a role of YAP/TAZ in mediating oncogenic effect of hypoxia.

Actin Cytoskeleton Integrates Upstream Signals

The actin cytoskeleton and Rho GTPases are not only important in maintaining cell morphology, but also play important roles in regulating cell proliferation and differentiation (Jaffe and Hall, 2005). Manipulation of the actin cytoskeleton, such as overexpression of Rho GTPases or inhibition of Rho by C3 toxin, dramatically modulates YAP/TAZ activity (Dupont et al., 2011; Yu et al., 2012; Zhao et al., 2012). Rho GTPases and changes in actin cytoskeleton dynamics have been demonstrated to be key mediators of mechanical cues, GPCR ligands, and cell attachment in regulating the Hippo pathway (Figure 3). Consistently, deletion of different regulators of the actin cytoskeleton impacts YAP and Yki activity. For instance, loss of actin-capping proteins or the *Capulet* gene (which inhibits actin polymerization) in *Drosophila* results in Yki activation and tissue overgrowth (Fernández et al., 2011; Sansores-Garcia et al., 2011). Similarly, knockdown of actin-capping proteins or filamentous actin (F-actin)-severing proteins (cofilin or gelsolin) in mammalian cells also leads to YAP activation (Aragona et al., 2013). In general, Rho GTPase activity and F-actin appear to activate YAP/TAZ, whereas destabilization of F-actin inhibits YAP/TAZ (Dupont et al., 2011; Miller et al., 2012; Wada et al., 2011; Yu et al., 2012; Zhao et al., 2012).

Spectrin proteins, in association with short actin filaments, are organized into an elastic polygonal meshwork that lines the intracellular side of the plasma membrane. In epithelial cells, localization of the spectrin network is usually polarized and present in both apical and basolateral domains. In addition to a supporting role for cell structure, the spectrin network may transmit diverse signals from cell microenvironment to regulate cellular functions (Bennett and Gilligan, 1993). Recently, three independent studies revealed a regulatory role of the spectrin network on the Hippo pathway, as disruption of the spectrin network in multiple *Drosophila* tissues leads to activation of Yki and tissue outgrowth (Deng et al., 2015; Fletcher et al., 2015; Wong et al., 2015). A similar phenomenon was also observed in mammalian cells to support a role of spectrin in YAP regulation (Figures 2 and 3). Therefore, spectrins or associated actin filaments may function as a major node for integrating upstream signals, such as mechanical cues.

Despite its apparent importance, it remains unclear how the actin cytoskeleton regulates activity of Hippo pathway core kinases. One possibility is that multiple Hippo pathway components are enriched at the apical domain via an actin-mediated mechanism that facilitates signal transduction, and actin remodeling may reinforce or disrupt the clustering of Hippo pathway components. In another scenario, the Hippo pathway could be regulated by contractile actomyosin and cellular tension (Figure 3). Inhibition of tension-related enzymes, such as nonmuscle myosin (by Blebbistatin), Rho kinases (ROCK, by Y27632), and myosin light-chain kinase (by ML-7), results in YAP/TAZ inhibition (Dupont et al., 2011; Wada et al., 2011). However, how tension is sensed by the Hippo pathway kinases is not fully understood. Treating cells with small molecules to inhibit cellular tension may also affect actin dynamics, so it is difficult to separate the effects of actin remodeling and tension generated by actomyosin. However, it has been proposed that, in *Drosophila*, Ajuba (Jub) plays a critical role in tension-induced Yki regulation. In response to high tension, Jub recruits Wts to intercellular junctions via interactions with α-catenin, thereby inhibiting Wts activity (Rauskolb et al., 2014). In addition, JNK activation upon mechanical strain has also been shown to repress LATS1 (Codelia et al., 2014).

Although it has been initially reported that YAP regulation by mechanical signals is independent of LATS1/2, recent studies suggest that LATS1/2 are involved in YAP/TAZ regulation by the actin cytoskeleton (Miller et al., 2012; Wada et al., 2011; Yu et al., 2012; Zhao et al., 2007, 2012). This discrepancy may be due to incomplete LATS1/2 depletion in knockdown experiments reported in earlier studies. Supporting a role for LATS1/2, the phosphorylation status and in vitro kinase activity of LATS1/2 are clearly regulated upon actin cytoskeleton rearrangement, and the kinetics are similar to that of YAP/TAZ phosphorylation (Yu et al., 2012; Zhao et al., 2012). The regulation of Yki/YAP/TAZ by spectrins or cell attachment/detachment has been shown to be dependent on Wts and LATS1/2 (Deng et al., 2015; Fletcher et al., 2015; Zhao et al., 2012). Moreover, during the preimplantation stage of mouse embryo development, LATS1/2 are essential for YAP regulation by small molecules targeting the actin cytoskeleton (Kono et al., 2014). Collectively, mechanical signals most likely act through Wts and LATS1/2 to influence the activity and function of Yki and YAP/TAZ.

Negative Feedback Regulation and Crosstalk

High Yki or YAP/TAZ activity, especially over the long term, results in tissue overgrowth or cancer (see below). Thus, the physiological fluctuation of Hippo signaling must be tightly controlled to avoid detrimental effects. In *Drosophila*, regulation of the Hippo pathway is fine-tuned by a built-in negative-feedback loop in which activation of Yki turns on the expression of upstream regulators, including Four-jointed, Ex, Mer, Kibra, and Wts (Cho et al., 2006; Genevet et al., 2010; Hamaratoglu et al., 2006; Jukam et al., 2013). Consistently, a similar negative feedback mechanism also exists in mammalian cells. YAP/TAZ directly induce the transcription of *NF2*, *LATS2*, and probably *MST1*, leading to LATS1/2 activation and YAP/TAZ inhibition (Chen et al., 2015b; Dai et al., 2015; Moroishi et al., 2015b). In addition, YAP/TAZ induce expression of angiomotin-like protein 2 (AMOTL2), a negative regulator of YAP (Mohseni et al., 2014; Zhao et al., 2008). This negative feedback loop is critical for maintaining the proper transient activation of YAP/TAZ upon stimulation. By these mechanisms, YAP and TAZ antagonize each other and may provide a buffer for fluctuations in Hippo pathway activity to ensure tissue homeostasis. When this negative feedback is disrupted, dysregulation of the Hippo pathway may lead to tumorigenesis.

Elucidating the Hippo pathway regulation and function is complicated by crosstalk between the Hippo pathway and many other developmental signaling pathways (reviewed in Varelas, 2014). Besides Wnt, YAP/TAZ may also be regulated by sonic hedgehog (SHH) signaling (Fernandez-L et al., 2009). On the other hand, YAP/TAZ has been shown to regulate the expression of ligands for Wnt, Shh, transforming growth factor β (TGF-β), JAK-STAT, EGFR, and Notch pathways (reviewed in Yu et al., 2015). The extensive signaling crosstalk may create a microenvironment rich in different factors, which in turn regulates cell fate through autocrine or paracrine mechanisms in both cell-autonomous and cell-non-autonomous manners.

Hippo Pathway in Early Embryonic Development

YAP/TAZ are critical during early embryonic development. Although TAZ knockout mice are viable, YAP knockout mice die at E8.5, and blastomeres stop dividing before the morula (16–32 cells) stage when YAP and TAZ are both deleted (Hossain et al., 2007; Makita et al., 2008; Morin-Kensicki et al., 2006; Nishioka et al., 2009; Tian et al., 2007). Therefore, the role of YAP/TAZ in early development is partially overlapping.

The first cell fate specification during embryogenesis occurs during preimplantation stage, in which the trophectoderm (TE) and inner cell mass (ICM) are formed. The TE consists of the outer cells of the blastocyst and forms extraembryonic tissues, while the ICM contains the inner cells of the blastocyst and give rise to the embryo proper and other tissues. The formation of the TE and ICM is mainly due to the position or polarity of cells in the morula, in which the inner and apolar cells form the ICM, while the outer and polar cells give rise to the TE (Sasaki, 2015). As early as the 16-cell stage, YAP/TAZ already show differential subcellular localization between inner and outer cells, and this difference lasts to the blastocyst stage. The different distribution of YAP/TAZ in the TE and ICM results in different gene expression signatures, especially the induction of TE-specific genes, such as *Cdx2* in outer cells, thus directing cell fate specification (Nishioka et al., 2009). Indeed, mouse embryos with TEAD4 knockout failed to develop TE cells, with all cells differentiating into ICM. On the other hand, depleting LATS1/2, NF2, or AMOT/AMOTL2 turns all cells into TE lineage, and these embryos failed to develop ICM-derived tissues (Cockburn et al., 2013; Hirate et al., 2013; Lorthongpanich et al., 2013). These results suggest that the Hippo pathway plays a key role in early embryonic cell specification.

Hippo Signaling in Organ Size Control and Tissue Homeostasis

The effect on organ size is the best-known physiological function of the Hippo pathway. In *Drosophila*, mutation of Hippo pathway kinases (*hpo* and *wts*) or upstream regulators (*ex*, *mer*, *kibra*, *ft*, etc.) leads to overgrowth of organs such as eyes, wings, or other appendages, and transgenic expression of *yki* results in a similar phenotype (reviewed in Halder and Johnson, 2011; Pan, 2010). The increased tissue/organ size is mainly due to Yki-induced cell proliferation and survival.

The effect of the Hippo pathway on organ size is highly conserved in mammals, as revealed by many studies performed in mice (summarized in Table 2). For instance, liver-specific transgenic *Yap* expression in mice can produce a dramatically enlarged liver. Remarkably, the liver returns to its normal size via apoptosis once *Yap* overexpression is turned off (Camargo et al., 2007; Dong et al., 2007). Similarly, liver-specific knockout of *Mst1/2*, *Sav1*, or *Nf2* also results in liver enlargement (Yin et al., 2013; Zhang et al., 2010; Zhou et al., 2009). The mouse embryonic heart is enlarged when *Sav1*, *Mst1/2*, or *Lats1/2* is deleted, and proliferation or apoptosis of cardiomyocytes is sensitive to genetic manipulation of *Yap* (Del Re et al., 2013; Heallen et al., 2011; Lin et al., 2014; von Gise et al., 2012; Xin et al., 2013; Xin et al., 2011).

However, not all organs are equally sensitive to Hippo pathway mutations. For instance, *Mst1/2* knockout results in dramatic overgrowth of liver, heart, stomach, and spleen, but not of kidney, lung, or limbs (Song et al., 2010). A possible explanation is that there are MST1/2-independent regulators of YAP/TAZ in these tissues. In the breast and intestine, tissue-specific deletion of *Yap* does not result in any defects in tissue structure or size (Cai et al., 2010; Chen et al., 2014). *Yap* knockout in the mouse liver leads to bile duct defects, but not a general reduction of liver size. This lack of effect could be due to the presence of TAZ, which should be activated due to the loss of feedback inhibition upon *Yap* deletion. Alternatively, the Hippo pathway, or YAP/TAZ activity, may be negligible for the size control of some organs. Nevertheless, the organ-specific effects of the Hippo pathway in size control may suggest that different principles are utilized for size regulation in different organs. For example, organ size could be determined by proliferation of differentiated cells or the number of progenitor cells in a pre-allocated pool.

Physiological signals upstream of the Hippo pathway important in organ size determination have yet to be identified. Mechanical force or tension may change as a result of organ growth and may inhibit YAP/TAZ when the organ reaches its final size. Alternatively, organ size may be restricted/induced by a soluble factor via autocrine/paracrine mechanisms, and the

Table 2. Physiological and Pathological Functions of Hippo Pathway Genes in Mice

Organ	Phenotypes	References
Liver	*Yap*-inducible expression results in hepatomegaly in a reversible manner and, in long term, leads to development of liver tumors.	Camargo et al., 2007; Dong et al., 2007
	Nf2, *Sav1*, or *Mst1/2* deletion causes hepatomegaly and results in hepatocellular carcinoma, cholangiocarcinoma, or bile duct hamartoma. *Mob1a/b* deletion also causes liver cancer.	Benhamouche et al., 2010; Lee et al., 2010; Lu et al., 2010; McClatchey et al., 1998; Nishio et al., 2012; Song et al., 2010; Yin et al., 2013; Zhang et al., 2010; Zhou et al., 2009
	Yap deletion causes bile duct defect, and YAP activity is involved in liver regeneration upon tissue damage.	Bai et al., 2012; Su et al., 2015; Yimlamai et al., 2014; Zhang et al., 2010
	Lkb1-deficiency-induced liver overgrowth is dependent on YAP activation.	Mohseni et al., 2014
Intestine	*Yap* transgenic expression causes intestinal dysplasia.	Camargo et al., 2007
	Deletion of *Sav1* or *Mst1/2* in mouse intestine results in expansion of progenitor cells, colonic polyps, or adenoma.	Cai et al., 2010; Lee et al., 2008; Zhou et al., 2011
	Deletion of *Yap* in mouse intestine shows no obvious phenotype but affects regeneration upon tissue damage.	Cai et al., 2010
	Yap transgenic expression (intestine specific) results in rapid loss of intestinal crypts by repressing Wnt signaling.	Barry et al., 2013
	Apc-deletion-induced expansion of intestinal crypts in a YAP/TAZ-dependent manner.	Azzolin et al., 2014; Cai et al., 2015
Skin	Deletion of *Sav1* or transgenic expression of *Yap* leads to expansion of basal progenitor cells and skin thickening; long-term activation of YAP results in squamous cell carcinoma. *Mob1a/b* deletion also causes skin cancer.	Camargo et al., 2007; Lee et al., 2008; Nishio et al., 2012
	Deletion of *Mst1/2* shows no clear phenotype. Deletion of *Ctnna1* (α-Catenin) leads to keratinocyte hyperproliferation and squamous cell carcinoma likely mediated by YAP activation.	Schlegelmilch et al., 2011
	Gnas (Gs) deletion causes basal-cell carcinoma partially dependent on YAP activation.	Iglesias-Bartolome et al., 2015
Heart	Deletion of *Sav1*, *Mst1/2*, or *Lats2* at embryonic stage or transgenic expression of active *Yap* mutant results in hyperproliferation of cardiomyocytes and enlargement of heart.	Del Re et al., 2013; Heallen et al., 2011; Lin et al., 2014; von Gise et al., 2012; Xin et al., 2013; Xin et al., 2011
	Deletion of *Yap* leads to heart hypoplasia, and more severe phenotype is observed when both *Yap* and *Taz* are deleted.	von Gise et al., 2012; Xin et al., 2011
	In adult heart, high YAP/TAZ activity enhances heart regeneration following cardiac damages such as myocardial infarction.	Heallen et al., 2013; Lin et al., 2014; Xin et al., 2013
Kidney	*Taz* deletion causes polycystic kidney disease, whereas *Yap* deletion leads to hypoplastic kidneys with severe defect in nephron morphogenesis.	Hossain et al., 2007; Makita et al., 2008; Reginensi et al., 2013; Tian et al., 2007
	Kidney with *Mst1/2* or *Sav1* deletion appears normal.	Reginensi et al., 2013; Song et al., 2010
Lung	*Taz*-deleted mice display abnormal alveolar structures.	Makita et al., 2008; Mitani et al., 2009; Tian et al., 2007
	Mst1/2 deletion leads to disrupted lung structures and neonatal lethality, which is dependent on high YAP/TAZ activity. *Mst1/2* deletion in adult lung bronchiolar epithelial cells results in airway hyperplasia and altered differentiation. YAP appears critical in regulating proximal-distal patterning of the lung, and a decrease in YAP activity ensures epithelial cells differentiation.	Lange et al., 2015; Lin et al., 2015a; Mahoney et al., 2014
	Mob1a/b deletion causes lung tumor.	Nishio et al., 2012
	Forkhead box A2 (FOXA2) has also been shown critical in mediating the effect of *Mst1/2* in lung development.	Chung et al., 2013

Table 2. *Continued*

Organ	Phenotypes	References
Pancreas	The effect of Hippo pathway is mainly in the exocrine compartment of the pancreas. During postnatal stage, deletion of *Mst1/2* increases the ratio of ductal and acinar cells and leads to pancreatitis-like autodigestion and a reduced size of pancreas.	Gao et al., 2013; George et al., 2012
	Kras (K12D)-mutant-induced pancreatic ductal adenocarcinoma requires YAP activity.	Zhang et al., 2014a
Nervous System	*Nf2* deletion causes an expansion of the neural progenitor pool and results in enlargement of the cortical hem, malformation of hippocampus (at late embryogenesis), and thickening of the neurocortex. *Nf2* deletion also affects development of corpus callosum, in which Yap-mediated overexpression of SLIT2 disrupts callosal axon pathfinding.	Lavado et al., 2013; Lavado et al., 2014
Mammary glands	*Yap* and *Sav1* are dispensable in mammary glands development. During pregnancy, *Yap* deletion results in hypoplasia and reduced alveolar structures; on the other hand, *Sav1* deletion or transgenic expression of *Yap* prevents terminal differentiation of mammary cells.	Chen et al., 2014
	Mob1a/b deletion causes breast tumor.	Nishio et al., 2012
	Yap deletion delays mammary tumor growth induced by polyoma middle T antigen (PyMT).	Chen et al., 2014
Muscle	*Yap* overexpression promotes proliferation of satellite cells and represses their differentiation. YAP downregulation reduces basal skeletal muscle fiber size, and YAP activity is required to relieve neurogenic muscle atrophy following injuries.	Judson et al., 2012; Watt et al., 2015

concentration of this factor is controlled by organ size. These two models are not mutually exclusive, and further investigations are needed to address how the Hippo pathway senses physiological cues to modulate organ size.

High YAP/TAZ activity has been observed in the stem or progenitor cells of multiple tissues, suggesting a role for YAP/TAZ in stem cell maintenance. For example, YAP is highly nuclear in basal progenitor cells and in intestinal stem cells localized at the crypt base (Barry et al., 2013; Camargo et al., 2007; Schlegel-milch et al., 2011; Zhang et al., 2011). Activation of YAP, either by transgenic expression of YAP or deletion of upstream regulators, usually results in expansion of progenitor cells, impaired cell differentiation, and hyperplasia of target tissues such as intestine, liver, skin, and nervous system (Cai et al., 2010; Camargo et al., 2007; Cao et al., 2008; Lee et al., 2008, 2010; Lu et al., 2010; Zhou et al., 2011).

The role of YAP/TAZ on cell proliferation and stem cell expansion suggests a critical function of YAP in normal tissue development and homeostasis. Indeed, tissue-specific deletion of *Yap* results in abnormalities of the heart, skin, and kidney (Reginensi et al., 2013; Schlegelmilch et al., 2011; von Gise et al., 2012; Xin et al., 2011). However, mammary glands and the intestine remain relatively normal upon *Yap* deletion (Cai et al., 2010; Chen et al., 2014; Zhou et al., 2011). These findings suggest that YAP is required for development and homeostasis of some, but not all, tissues in mice. In human, *TEAD1* mutations are found in Sveinsson chorioretinal atrophy, a disease characterized by chorioretinal degeneration, and Aicardi syndrome, a congenital neurodevelopmental disorder (Fossdal et al., 2004; Schrauwen et al., 2015). In addition, loss-of-function mutations of *YAP*

have been identified in both isolated and syndromic optic fissure closure defects (Williamson et al., 2014). Hence, the loss of TEAD-mediated YAP transcriptional activity plays a role in some degeneration-related disorders in humans.

Even though it is not required for development and normal homeostasis of some tissues, YAP activity is critical for tissue regeneration upon certain types of damage. For example, *Yap* deletion severely compromises pregnancy-induced mammary tissue growth, although virgin mammary development was normal (Chen et al., 2014). Likewise, in wild-type mice, the intestines can effectively regenerate following colitis induced by dextran sulfate sodium (DSS) treatment; however, the regenerative capability is severely impeded in conditional *Yap* knockout mice (Cai et al., 2010). Similar results have also been observed in *Drosophila* midgut regeneration upon DSS-induced injury (Karpowicz et al., 2010; Ren et al., 2010; Shaw et al., 2010). Normally the liver regenerates efficiently following liver damage. For instance, after partial hepatectomy, hepatocytes start to proliferate to restore liver mass in a few days, and in this process, YAP activity is induced and is most likely required for complete liver regeneration (Grijalva et al., 2014; Su et al., 2015; Wu et al., 2013; Yimlamai et al., 2014). In contrast to the intestine and liver, tissue regeneration of the adult heart is very limited. However, inactivation of the Hippo pathway or transgenic expression of *Yap* restores some myocardial regenerative capability, although the efficiency is low. In contrast, cardiac-specific deletion of *Yap* impedes regeneration of the neonatal heart (Heallen et al., 2013; Lin et al., 2014; Xin et al., 2013). Taken together, these results indicate that YAP plays a significant role in regeneration of multiple tissues.

Hippo Signaling in Cancer

Long-term YAP activation, such as transgenic expression of *Yap* in the mouse liver, results in cell transformation and tumor development (Dong et al., 2007), indicating the power of the Hippo pathway in cancer initiation and progression. Evidence of the Hippo pathway in tumorigenesis based on mouse models is summarized in Table 2, which generally supports an oncogenic role for YAP/TAZ, as well as a tumor-suppressive function for Hippo pathway upstream components.

The tumor-promoting activity of YAP is largely dependent on a TEAD-mediated transcription program, as YAP-induced liver cancer is fully blocked by expression of a dominant-negative TEAD that is able to sequester both YAP and TAZ (Liu-Chittenden et al., 2012). At the cellular level, YAP activation is important for cell proliferation, survival, migration, and invasion. High YAP or TAZ activity enables the cell to escape contact inhibition and anoikis and to support anchorage-independent growth (Chan et al., 2008; Zhao et al., 2007, 2012). YAP induces expression of ZEB1/2 to stimulate the epithelial-to-mesenchymal transition (EMT), which is a key step for tumor metastasis (Gao et al., 2014; Lei et al., 2008; Liu et al., 2010b; Overholtzer et al., 2006). In addition, YAP is able to promote genomic instability (Fernandez-L et al., 2012), and TAZ is required to sustain self-renewal and the tumor-initiation capacity of breast cancer stem cells (Cordenonsi et al., 2011).

Accumulating evidence suggests that the Hippo pathway is dysregulated in many human cancers. Elevated YAP/TAZ expression or nuclear enrichment of YAP/TAZ has been observed in many types of cancers, including liver, breast, lung, colon, ovary, and others (Chan et al., 2008; Moroishi et al., 2015a; Steinhardt et al., 2008). However, the majority of cancers with high YAP/TAZ activity have not been linked to genetic mutations of the Hippo pathway, and the overall genetic alteration rate of Hippo pathway components in human cancer is relatively low (Table 3).

One well-characterized example of a Hippo pathway mutation associated with cancer is in *NF2*, which causes neurofibromatosis 2 lesions, including schwannomas and meningiomas (Xiao et al., 2003). Moreover, inactivating *NF2* mutations are also observed in 40%–50% of malignant mesothelioma (Sekido, 2011). Importantly, even heterozygous deletion of *Yap* completely blocks liver tumorigenesis induced by *Nf2* knockout in mice, indicating that YAP activation is the major mechanism mediating the tumorigenic potential of *Nf2* mutations (Zhang et al., 2010).

YAP gene amplification may contribute to a portion of hepatocellular carcinomas, medulloblastomas, and esophageal squamous cell carcinomas (Fernandez-L et al., 2009; Overholtzer et al., 2006; Song et al., 2014; Zender et al., 2006). Gene fusions involving *YAP* or *TAZ* have also been discovered in human cancers. Remarkably, virtually all epithelioid hemangioendotheliomas contain gene fusions of *TAZ-CAMTA1*, *TAZ-FOSB*, or *YAP-TFE3* (Antonescu et al., 2013; Errani et al., 2011; Flucke et al., 2014; Tanas et al., 2011). In addition, *YAP* gene fusions with MAMLD1 or *C11orf95* have been discovered in a subset of ependymal tumors (Pajtler et al., 2015; Parker et al., 2014). It is worth noting that, in both epithelioid hemangioendotheliomas and ependymal tumors, all YAP/TAZ fusion proteins retain their N-terminal TEAD-binding domain but lose the C-terminal trans-

activation domain. These observations suggest that these YAP fusions may still bind to and activate the TEAD-dependent transcriptional program to promote tumorigenesis. Indeed, neural stem cells carrying the *YAP-C11orf95* fusion gene can effectively form brain tumors when grafted into mice (Parker et al., 2014). In addition, a familial *YAP* point mutation (R331W) has also been reported to correlate with a high incidence of lung adenocarcinomas (Chen et al., 2015a).

Aberrant GPCR signaling often results in tumorigenesis, so it is possible that GPCR dysregulation can cause cancer by activating YAP/TAZ. *GNAQ*- or *GNA11*- (encoding Gα_q or Gα_{11}, respectively) activating mutations have been identified in ∼80% of uveal melanomas and function as driver mutations (Van Raamsdonk et al., 2009; Van Raamsdonk et al., 2010). Recent studies showed that YAP is constitutively activated in *GNAQ*- or *GNA11*-mutated uveal melanomas, and the high YAP activity contributes to tumor growth (Feng et al., 2014; Yu et al., 2014). In mice, deletion of *Gnas* (encoding Gα_S) in skin stem cells initiates basal-cell carcinogenesis, which is partially dependent on YAP (Iglesias-Bartolome et al., 2015). Moreover, expression of a viral GPCR induces tumorigenesis in Kaposi's sarcoma, where YAP/TAZ also play a critical role (Liu et al., 2015).

LATS1/2 mutations or gene fusion have been sporadically identified in different cancers, which may lead to YAP/TAZ activation (Table 3). In addition, crosstalk with other cancer-related signaling pathways also likely contributes to high YAP/TAZ activity in cancers that have no mutations of Hippo pathway components. For example, *KRAS*, *APC*, and *LKB1* mutations have all been reported to activate YAP/TAZ (Azzolin et al., 2014; Gao et al., 2014; Mohseni et al., 2014; Zhang et al., 2014a).

YAP/TAZ activity is also linked to drug resistance and cancer relapse. Cultured breast cancer cells with high YAP/TAZ activity show resistance to drugs such as taxol, 5-fluorouracil, and doxorubicin (Cordenonsi et al., 2011; Lai et al., 2011; Touil et al., 2014). Furthermore, lung and colon cancer cells with high YAP activity are resistant to RAF- and MEK-targeted therapies (Lin et al., 2015b). Tamoxifen is commonly used to treat estrogen-receptor- (ER, a nuclear receptor) positive breast cancer; however, some ER-positive breast cancers are insensitive to tamoxifen. Recently, tamoxifen has been shown to activate YAP/TAZ by stimulating the membrane estrogen receptor GPER, a GPCR. Therefore, activation of GPER by tamoxifen or estrogen may contribute to tumor growth and drug resistance (Zhou et al., 2015). Amplification of the *YAP* gene has been associated with cancer relapse in *KRAS*-driven colon and pancreatic cancers (Kapoor et al., 2014; Shao et al., 2014). Thus, inhibition of YAP/TAZ will not only target tumor initiation and progression, but also potentially sensitize cancer cells to chemotherapies and prevent cancer relapse.

Notably, in contrast to its oncogenic function in most solid tumors, YAP seems to play a tumor-suppressor role in hematological cancers. The *YAP* gene locus is frequently deleted in hematological cancer, and expression of YAP or inhibition of MST1 leads to growth inhibition and increased apoptosis (Cottini et al., 2014). Currently, the underlying mechanism responsible for this tumor-suppressor function of YAP in hematological cancers is not well understood.

Table 3. Genetic Alterations of Hippo Pathway Genes in Human Cancers

Gene	Alteration	Cancer type	References
NF2	mutation or deletion	mesothelioma neurofibromatosis type 2 (schwannoma, meningioma)	Sekido, 2011 Rouleau et al., 1993
LATS1/2	gene fusion (*LATS1-PSEN1*)	mesothelioma	Miyanaga et al., 2015
	LATS2 deletion	mesothelioma	Murakami et al., 2011
	LATS1/2 mutations	sporadic in different cancers	Yu et al., 2013b
YAP	amplification	hepatocellular carcinoma medulloblastoma esophageal squamous cell carcinoma	Fernandez-L et al., 2009 Overholtzer et al., 2006 Song et al., 2014 Zender et al., 2006
	mutation (R331W)	lung adenocarcinoma	Chen et al., 2015a
	gene fusion (*YAP-TFE3*, *YAP-ESR1*, *YAP-C11orf95*, and *YAP-MAMLD1*)	epithelioid hemangioendothelioma luminal breast cancer ependymal tumors	Antonescu et al., 2013 Flucke et al., 2014 Li et al., 2013 Pajtler et al., 2015 Parker et al., 2014
	deletion	hematological cancer	Cottini et al., 2014
TAZ	gene fusion (*TAZ-CAMTA1* and *TAZ-FOSB*)	epithelioid hemangioendothelioma	Errani et al., 2011 Flucke et al., 2014 Tanas et al., 2011
GNAQ/GNA11	activating mutation	uveal melanoma	Van Raamsdonk et al., 2009 Van Raamsdonk et al., 2010

Therapeutic Targeting of Hippo Signaling

The core Hippo pathway is a kinase cascade, and protein kinases are usually druggable. Thus, inhibitors for MAP4K4, MST1/2, or LATS1/2 may be developed to induce YAP/TAZ activity and facilitate the process of wound healing, tissue repair, or regeneration and possibly for treating degenerative diseases (Figure 4). For example, temporal inhibition of MST1/2 or LATS1/2 may promote myocardial regeneration or survival that would be beneficial for heart attack patients. It is also possible that inhibitors of MST1/2 or LATS1/2 could be used for treating hematological cancers.

Generally, MST1/2 and LATS1/2 are tumor suppressors, and inhibition of MST1/2 or LATS1/2 may promote tumor growth in most instances. On the other hand, inhibiting YAP/TAZ activity would offer a new and attractive anti-cancer strategy (Park and Guan, 2013). The function of YAP/TAZ is primarily mediated by TEADs, so small molecules disrupting the YAP/TAZ-TEAD interaction will function as YAP/TAZ inhibitors. Indeed, porphyrin family molecules, especially verteporfin, are able to disrupt the interaction between YAP/TAZ and TEADs, and verteporfin can block transcription of YAP/TAZ target genes and suppress liver overgrowth induced by YAP overexpression or NF2 inactivation in mice (Liu-Chittenden et al., 2012). However, verteporfin has general cellular toxicity and low aqueous solubility. Based on structural information from the YAP-TEAD and VGLL4-TEAD complex, a polypeptide termed "super-TDU" has been designed to block YAP-TEAD interaction and has been shown to suppress tumor growth in mouse models (Jiao et al., 2014).

It is challenging to design direct activators for protein kinases. However, LATS1/2 may be activated indirectly by molecules targeting their upstream regulators. The very first small molecule (dobutamine) identified with an inhibitory effect on YAP is an antagonist of a GPCR receptor (Bao et al., 2011). Since then, many indirect inhibitors for YAP/TAZ have been identified, including phosphodiesterase inhibitors rolipram and ibudilast (Yu et al., 2013a). The Rho family GTPases have a strong inhibitory effect on LATS1/2, and membrane localization is important for Rho cellular function. Indeed, mevalonate metabolic pathway inhibitor statins can block membrane translocation of Rho GTPases and indirectly inhibit YAP/TAZ activity (Mi et al., 2015; Sorrentino et al., 2014; Wang et al., 2014). It will be interesting to test whether these drugs are effective in suppressing tumor growth in mouse models and to perform epidemiologic studies to determine whether patients using statins or rolipram have a lower incidence of cancer. Given the important function of the Hippo pathway in regulating cell proliferation and tissue homeostasis, it represents an exciting and previously unexplored field for cancer therapy.

Outstanding Questions

Despite the rapid research progress in the Hippo pathway, some key questions remain unanswered, and new questions are emerging. Listed below are some of the key questions in the Hippo field.

(1) What are the molecular mechanisms regulating MST1/2 activity? Many upstream signals have been convincingly shown to regulate LATS1/2 phosphorylation and kinase activity. However, neither MST1/2 phosphorylation nor its kinase activity is strongly modulated by upstream
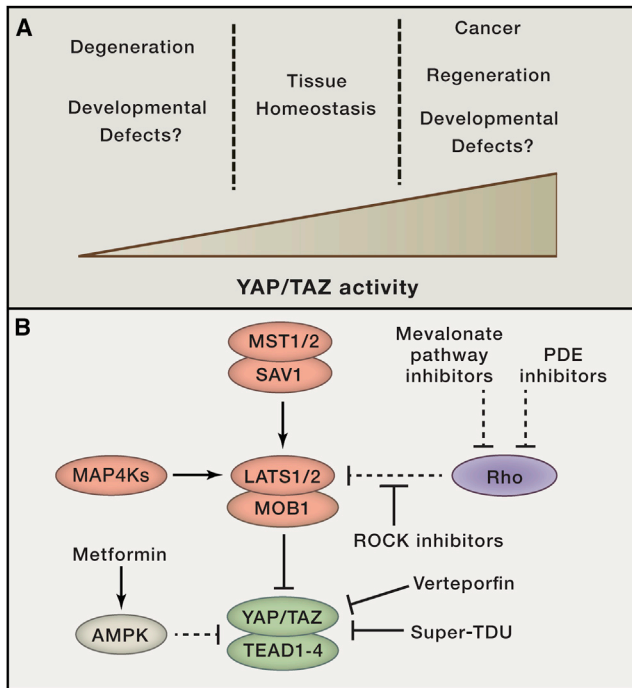
Figure 4. Therapeutic Targeting of the Hippo Pathway
(A) Potential roles of YAP/TAZ activity in tissue development and diseases. A confined window of YAP/TAZ activity is required for normal tissue development and homeostasis.
(B) Strategies for targeting YAP/TAZ activity. Inhibitors for MST1/2, MAP4K4, and LATS1/2 can activate YAP/TAZ. YAP/TAZ-TEAD interaction may be disrupted by small molecules directly (Verteporfin) or AMPK activators (Metformin). Small molecules inhibiting Rho-family GTPases or ROCK can indirectly activate LATS1/2, leading to YAP/TAZ inhibition.

signals. *Drosophila* misshapen (a member of MAP4K in mammals) acts upstream of Wts. An interesting question is whether MST1/2 and MAP4Ks mediate different or similar upstream signals to activate LATS1/2.

(2) Where are Hippo pathway components localized in mammalian cells? This may be key to understanding how the Hippo pathway is regulated in response to upstream signals. It is obvious that phosphorylated YAP/TAZ are enriched in the cytoplasm and dephosphorylated YAP/TAZ in the nucleus. However, it is less clear where YAP/TAZ phosphorylation and dephosphorylation occur. A related question is what are the mechanisms underlying YAP/TAZ translocation between the nucleus and cytoplasm?

(3) How are LATS1/2 regulated by actin remodeling and/or cellular tension? This is one of the key questions in understanding the biochemical mechanism of the Hippo kinase cascade regulation. Accumulating evidence suggests that the actin cytoskeleton and cellular tension play a key role in LATS1/2 regulation and appear to act downstream of many, if not most, upstream signals. The actin cytoskeleton and cellular tension are intertwined, thus the question is: which one plays a more direct role in regulating Hippo core components?

(4) What is the mechanism of organ size sensing? Although many signals are reported to regulate the Hippo pathway in vitro, so far none have been demonstrated to play a key role in organ size control in vivo. Uncovering this magic signal will solve a key question in developmental biology.

(5) How does YAP become deregulated in cancer? It is clear that YAP/TAZ activation is observed in a broad spectrum of human cancers, although mutations in Hippo pathway genes are rare. This interesting conundrum indicates that the Hippo pathway may be regulated broadly by many other cancer-driving pathways.

REFERENCES

Antonescu, C.R., Le Loarer, F., Mosquera, J.M., Sboner, A., Zhang, L., Chen, C.L., Chen, H.W., Pathan, N., Krausz, T., Dickson, B.C., et al. (2013). Novel YAP1-TFE3 fusion defines a distinct subset of epithelioid hemangioendothelioma. Genes Chromosomes Cancer *52*, 775–784.

Aragona, M., Panciera, T., Manfrin, A., Giulitti, S., Michielin, F., Elvassore, N., Dupont, S., and Piccolo, S. (2013). A mechanical checkpoint controls multicellular growth through YAP/TAZ regulation by actin-processing factors. Cell *154*, 1047–1059.

Azzolin, L., Zanconato, F., Bresolin, S., Forcato, M., Basso, G., Bicciato, S., Cordenonsi, M., and Piccolo, S. (2012). Role of TAZ as mediator of Wnt signaling. Cell *151*, 1443–1456.

Azzolin, L., Panciera, T., Soligo, S., Enzo, E., Bicciato, S., Dupont, S., Bresolin, S., Frasson, C., Basso, G., Guzzardo, V., et al. (2014). YAP/TAZ incorporation in the β-catenin destruction complex orchestrates the Wnt response. Cell *158*, 157–170.

Bai, H., Zhang, N., Xu, Y., Chen, Q., Khan, M., Potter, J.J., Nayar, S.K., Cornish, T., Alpini, G., Bronk, S., et al. (2012). Yes-associated protein regulates the hepatic response after bile duct ligation. Hepatology *56*, 1097–1107.

Bao, Y., Nakagawa, K., Yang, Z., Ikeda, M., Withanage, K., Ishigami-Yuasa, M., Okuno, Y., Hata, S., Nishina, H., and Hata, Y. (2011). A cell-based assay to screen stimulators of the Hippo pathway reveals the inhibitory effect of dobutamine on the YAP-dependent gene transcription. J. Biochem. *150*, 199–208.

Barry, E.R., Morikawa, T., Butler, B.L., Shrestha, K., de la Rosa, R., Yan, K.S., Fuchs, C.S., Magness, S.T., Smits, R., Ogino, S., et al. (2013). Restriction of intestinal stem cell expansion and the regenerative response by YAP. Nature *493*, 106–110.

Baumgartner, R., Poernbacher, I., Buser, N., Hafen, E., and Stocker, H. (2010). The WW domain protein Kibra acts upstream of Hippo in Drosophila. Dev. Cell *18*, 309–316.

Benham-Pyle, B.W., Pruitt, B.L., and Nelson, W.J. (2015). Cell adhesion. Mechanical strain induces E-cadherin-dependent Yap1 and β-catenin activation to drive cell cycle entry. Science *348*, 1024–1027.

Benhamouche, S., Curto, M., Saotome, I., Gladden, A.B., Liu, C.H., Giovannini, M., and McClatchey, A.I. (2010). Nf2/Merlin controls progenitor homeostasis and tumorigenesis in the liver. Genes Dev. *24*, 1718–1730.

Bennett, V., and Gilligan, D.M. (1993). The spectrin-based membrane skeleton and micron-scale organization of the plasma membrane. Annu. Rev. Cell Biol. *9*, 27–66.

Bennett, F.C., and Harvey, K.F. (2006). Fat cadherin modulates organ size in Drosophila via the Salvador/Warts/Hippo signaling pathway. Curr. Biol. *16*, 2101–2110.

Boggiano, J.C., Vanderzalm, P.J., and Fehon, R.G. (2011). Tao-1 phosphorylates Hippo/MST kinases to regulate the Hippo-Salvador-Warts tumor suppressor pathway. Dev. Cell *21*, 888–895.

Cai, J., Zhang, N., Zheng, Y., de Wilde, R.F., Maitra, A., and Pan, D. (2010). The Hippo signaling pathway restricts the oncogenic potential of an intestinal regeneration program. Genes Dev. *24*, 2383–2388.

Cai, J., Maitra, A., Anders, R.A., Taketo, M.M., and Pan, D. (2015). β-Catenin destruction complex-independent regulation of Hippo-YAP signaling by APC in intestinal tumorigenesis. Genes Dev. *29*, 1493–1506.

Callus, B.A., Verhagen, A.M., and Vaux, D.L. (2006). Association of mammalian sterile twenty kinases, Mst1 and Mst2, with hSalvador via C-terminal coiled-coil domains, leads to its stabilization and phosphorylation. FEBS J. *273*, 4264–4276.

Camargo, F.D., Gokhale, S., Johnnidis, J.B., Fu, D., Bell, G.W., Jaenisch, R., and Brummelkamp, T.R. (2007). YAP1 increases organ size and expands undifferentiated progenitor cells. Curr. Biol. *17*, 2054–2060.

Cao, X., Pfaff, S.L., and Gage, F.H. (2008). YAP regulates neural progenitor cell number via the TEA domain transcription factor. Genes Dev. *22*, 3320–3334.

Chan, E.H., Nousiainen, M., Chalamalasetty, R.B., Schäfer, A., Nigg, E.A., and Silljé, H.H. (2005). The Ste20-like kinase Mst2 activates the human large tumor suppressor kinase Lats1. Oncogene *24*, 2076–2086.

Chan, S.W., Lim, C.J., Guo, K., Ng, C.P., Lee, I., Hunziker, W., Zeng, Q., and Hong, W. (2008). A role for TAZ in migration, invasion, and tumorigenesis of breast cancer cells. Cancer Res. *68*, 2592–2598.

Chaulk, S.G., Lattanzi, V.J., Hiemer, S.E., Fahlman, R.P., and Varelas, X. (2014). The Hippo pathway effectors TAZ/YAP regulate dicer expression and microRNA biogenesis through Let-7. J. Biol. Chem. *289*, 1886–1891.

Chen, C.L., Gajewski, K.M., Hamaratoglu, F., Bossuyt, W., Sansores-Garcia, L., Tao, C., and Halder, G. (2010). The apical-basal cell polarity determinant Crumbs regulates Hippo signaling in Drosophila. Proc. Natl. Acad. Sci. USA *107*, 15810–15815.

Chen, Q., Zhang, N., Gray, R.S., Li, H., Ewald, A.J., Zahnow, C.A., and Pan, D. (2014). A temporal requirement for Hippo signaling in mammary gland differentiation, growth, and tumorigenesis. Genes Dev. *28*, 432–437.

Chen, H.Y., Yu, S.L., Ho, B.C., Su, K.Y., Hsu, Y.C., Chang, C.S., Li, Y.C., Yang, S.Y., Hsu, P.Y., Ho, H., et al. (2015a). R331W Missense Mutation of Oncogene YAP1 Is a Germline Risk Allele for Lung Adenocarcinoma With Medical Actionability. J. Clin. Oncol. *33*, 2303–2310.

Chen, Q., Zhang, N., Xie, R., Wang, W., Cai, J., Choi, K.S., David, K.K., Huang, B., Yabuta, N., Nojima, H., et al. (2015b). Homeostatic control of Hippo signaling activity revealed by an endogenous activating mutation in YAP. Genes Dev. *29*, 1285–1297.

Cho, E., Feng, Y., Rauskolb, C., Maitra, S., Fehon, R., and Irvine, K.D. (2006). Delineation of a Fat tumor suppressor pathway. Nat. Genet. *38*, 1142–1150.

Chung, C., Kim, T., Kim, M., Kim, M., Song, H., Kim, T.S., Seo, E., Lee, S.H., Kim, H., Kim, S.K., et al. (2013). Hippo-Foxa2 signaling pathway plays a role in peripheral lung maturation and surfactant homeostasis. Proc. Natl. Acad. Sci. USA *110*, 7732–7737.

Cockburn, K., Biechele, S., Garner, J., and Rossant, J. (2013). The Hippo pathway member Nf2 is required for inner cell mass specification. Curr. Biol. *23*, 1195–1201.

Codelia, V.A., Sun, G., and Irvine, K.D. (2014). Regulation of YAP by mechanical strain through Jnk and Hippo signaling. Curr. Biol. *24*, 2012–2017.

Cordenonsi, M., Zanconato, F., Azzolin, L., Forcato, M., Rosato, A., Frasson, C., Inui, M., Montagner, M., Parenti, A.R., Poletti, A., et al. (2011). The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. Cell *147*, 759–772.

Cottini, F., Hideshima, T., Xu, C., Sattler, M., Dori, M., Agnelli, L., ten Hacken, E., Bertilaccio, M.T., Antonini, E., Neri, A., et al. (2014). Rescue of Hippo coactivator YAP1 triggers DNA damage-induced apoptosis in hematological cancers. Nat. Med. *20*, 599–606.

Dai, X., Liu, H., Shen, S., Guo, X., Yan, H., Ji, X., Li, L., Huang, J., Feng, X.H., and Zhao, B. (2015). YAP activates the Hippo pathway in a negative feedback loop. Cell Res. *25*, 1175–1178.

Del Re, D.P., Yang, Y., Nakano, N., Cho, J., Zhai, P., Yamamoto, T., Zhang, N., Yabuta, N., Nojima, H., Pan, D., and Sadoshima, J. (2013). Yes-associated protein isoform 1 (Yap1) promotes cardiomyocyte survival and growth to protect against myocardial ischemic injury. J. Biol. Chem. *288*, 3977–3988.

Deng, H., Wang, W., Yu, J., Zheng, Y., Qing, Y., and Pan, D. (2015). Spectrin regulates Hippo signaling by modulating cortical actomyosin activity. eLife *4*, e06567.

DeRan, M., Yang, J., Shen, C.H., Peters, E.C., Fitamant, J., Chan, P., Hsieh, M., Zhu, S., Asara, J.M., Zheng, B., et al. (2014). Energy stress regulates hippo-YAP signaling involving AMPK-mediated regulation of angiomotin-like 1 protein. Cell Rep. *9*, 495–503.

Dong, J., Feldmann, G., Huang, J., Wu, S., Zhang, N., Comerford, S.A., Gayyed, M.F., Anders, R.A., Maitra, A., and Pan, D. (2007). Elucidation of a universal size-control mechanism in Drosophila and mammals. Cell *130*, 1120–1133.

Dupont, S., Morsut, L., Aragona, M., Enzo, E., Giulitti, S., Cordenonsi, M., Zanconato, F., Le Digabel, J., Forcato, M., Bicciato, S., et al. (2011). Role of YAP/TAZ in mechanotransduction. Nature *474*, 179–183.

Enzo, E., Santinon, G., Pocaterra, A., Aragona, M., Bresolin, S., Forcato, M., Grifoni, D., Pession, A., Zanconato, F., Guzzo, G., et al. (2015). Aerobic glycolysis tunes YAP/TAZ transcriptional activity. EMBO J. *34*, 1349–1370.

Errani, C., Zhang, L., Sung, Y.S., Hajdu, M., Singer, S., Maki, R.G., Healey, J.H., and Antonescu, C.R. (2011). A novel WWTR1-CAMTA1 gene fusion is a consistent abnormality in epithelioid hemangioendothelioma of different anatomic sites. Genes Chromosomes Cancer *50*, 644–653.

Fan, R., Kim, N.G., and Gumbiner, B.M. (2013). Regulation of Hippo pathway by mitogenic growth factors via phosphoinositide 3-kinase and phosphoinositide-dependent kinase-1. Proc. Natl. Acad. Sci. USA *110*, 2569–2574.

Feng, X., Degese, M.S., Iglesias-Bartolome, R., Vaque, J.P., Molinolo, A.A., Rodrigues, M., Zaidi, M.R., Ksander, B.R., Merlino, G., Sodhi, A., et al. (2014). Hippo-independent activation of YAP by the GNAQ uveal melanoma oncogene through a trio-regulated rho GTPase signaling circuitry. Cancer Cell *25*, 831–845.

Fernández, B.G., Gaspar, P., Brás-Pereira, C., Jezowska, B., Rebelo, S.R., and Janody, F. (2011). Actin-Capping Protein and the Hippo pathway regulate F-actin and tissue growth in Drosophila. Development *138*, 2337–2346.

Fernandez-L, A., Northcott, P.A., Dalton, J., Fraga, C., Ellison, D., Angers, S., Taylor, M.D., and Kenney, A.M. (2009). YAP1 is amplified and up-regulated in hedgehog-associated medulloblastomas and mediates Sonic hedgehog-driven neural precursor proliferation. Genes Dev. *23*, 2729–2741.

Fernandez-L, A., Squatrito, M., Northcott, P., Awan, A., Holland, E.C., Taylor, M.D., Nahlé, Z., and Kenney, A.M. (2012). Oncogenic YAP promotes radioresistance and genomic instability in medulloblastoma through IGF2-mediated Akt activation. Oncogene *31*, 1923–1937.

Fletcher, G.C., Elbediwy, A., Khanal, I., Ribeiro, P.S., Tapon, N., and Thompson, B.J. (2015). The Spectrin cytoskeleton regulates the Hippo signalling pathway. EMBO J. *34*, 940–954.

Flucke, U., Vogels, R.J., de Saint Aubain Somerhausen, N., Creytens, D.H., Riedl, R.G., van Gorp, J.M., Milne, A.N., Huysentruyt, C.J., Verdijk, M.A., van Asseldonk, M.M., et al. (2014). Epithelioid Hemangioendothelioma: clinicopathologic, immunhistochemical, and molecular genetic analysis of 39 cases. Diagn. Pathol. *9*, 131.

Fossdal, R., Jonasson, F., Kristjansdottir, G.T., Kong, A., Stefansson, H., Gosh, S., Gulcher, J.R., and Stefansson, K. (2004). A novel TEAD1 mutation is the causative allele in Sveinsson's chorioretinal atrophy (helicoid peripapillary chorioretinal degeneration). Hum. Mol. Genet. *13*, 975–981.

Gailite, I., Aerne, B.L., and Tapon, N. (2015). Differential control of Yorkie activity by LKB1/AMPK and the Hippo/Warts cascade in the central nervous system. Proc. Natl. Acad. Sci. USA *112*, E5169–E5178.

Gao, T., Zhou, D., Yang, C., Singh, T., Penzo-Mendez, A., Maddipati, R., Tzatsos, A., Bardeesy, N., Avruch, J., and Stanger, B.Z. (2013). Hippo signaling regulates differentiation and maintenance in the exocrine pancreas. Gastroenterology *144*, 1543–1553.

Gao, Y., Zhang, W., Han, X., Li, F., Wang, X., Wang, R., Fang, Z., Tong, X., Yao, S., Li, F., et al. (2014). YAP inhibits squamous transdifferentiation of Lkb1-deficient lung adenocarcinoma through ZEB2-dependent DNp63 repression. Nat. Commun. *5*, 4629.

Genevet, A., Wehr, M.C., Brain, R., Thompson, B.J., and Tapon, N. (2010). Kibra is a regulator of the Salvador/Warts/Hippo signaling network. Dev. Cell *18*, 300–308.

George, N.M., Day, C.E., Boerner, B.P., Johnson, R.L., and Sarvetnick, N.E. (2012). Hippo signaling regulates pancreas development through inactivation of Yap. Mol. Cell. Biol. *32*, 5116–5128.

Gong, R., Hong, A.W., Plouffe, S.W., Zhao, B., Liu, G., Yu, F.X., Xu, Y., and Guan, K.L. (2015). Opposing roles of conventional and novel PKC isoforms in Hippo-YAP pathway regulation. Cell Res. *25*, 985–988.

Goulev, Y., Fauny, J.D., Gonzalez-Marti, B., Flagiello, D., Silber, J., and Zider, A. (2008). SCALLOPED interacts with YORKIE, the nuclear effector of the hippo tumor-suppressor pathway in Drosophila. Curr. Biol. *18*, 435–441.

Grijalva, J.L., Huizenga, M., Mueller, K., Rodriguez, S., Brazzo, J., Camargo, F., Sadri-Vakili, G., and Vakili, K. (2014). Dynamic alterations in Hippo signaling pathway and YAP activation during liver regeneration. Am. J. Physiol. Gastrointest. Liver Physiol. *307*, G196–G204.

Guo, T., Lu, Y., Li, P., Yin, M.X., Lv, D., Zhang, W., Wang, H., Zhou, Z., Ji, H., Zhao, Y., and Zhang, L. (2013). A novel partner of Scalloped regulates Hippo signaling via antagonizing Scalloped-Yorkie activity. Cell Res. *23*, 1201–1214.

Halder, G., and Johnson, R.L. (2011). Hippo signaling: growth control and beyond. Development *138*, 9–22.

Hamaratoglu, F., Willecke, M., Kango-Singh, M., Nolo, R., Hyun, E., Tao, C., Jafar-Nejad, H., and Halder, G. (2006). The tumour-suppressor genes NF2/Merlin and Expanded act through Hippo signalling to regulate cell proliferation and apoptosis. Nat. Cell Biol. *8*, 27–36.

Harvey, K.F., Pfleger, C.M., and Hariharan, I.K. (2003). The Drosophila Mst ortholog, hippo, restricts growth and cell proliferation and promotes apoptosis. Cell *114*, 457–467.

Heallen, T., Zhang, M., Wang, J., Bonilla-Claudio, M., Klysik, E., Johnson, R.L., and Martin, J.F. (2011). Hippo pathway inhibits Wnt signaling to restrain cardiomyocyte proliferation and heart size. Science *332*, 458–461.

Heallen, T., Morikawa, Y., Leach, J., Tao, G., Willerson, J.T., Johnson, R.L., and Martin, J.F. (2013). Hippo signaling impedes adult heart regeneration. Development *140*, 4683–4690.

Hirate, Y., Hirahara, S., Inoue, K., Suzuki, A., Alarcon, V.B., Akimoto, K., Hirai, T., Hara, T., Adachi, M., Chida, K., et al. (2013). Polarity-dependent distribution of angiomotin localizes Hippo signaling in preimplantation embryos. Curr. Biol. *23*, 1181–1194.

Hossain, Z., Ali, S.M., Ko, H.L., Xu, J., Ng, C.P., Guo, K., Qi, Z., Ponniah, S., Hong, W., and Hunziker, W. (2007). Glomerulocystic kidney disease in mice with a targeted inactivation of Wwtr1. Proc. Natl. Acad. Sci. USA *104*, 1631–1636.

Huang, J., Wu, S., Barrera, J., Matthews, K., and Pan, D. (2005). The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the Drosophila Homolog of YAP. Cell *122*, 421–434.

Huang, W., Lv, X., Liu, C., Zha, Z., Zhang, H., Jiang, Y., Xiong, Y., Lei, Q.Y., and Guan, K.L. (2012). The N-terminal phosphodegron targets TAZ/WWTR1 protein for SCFβ-TrCP-dependent degradation in response to phosphatidylinositol 3-kinase inhibition. J. Biol. Chem. *287*, 26245–26253.

Iglesias-Bartolome, R., Torres, D., Marone, R., Feng, X., Martin, D., Simaan, M., Chen, M., Weinstein, L.S., Taylor, S.S., Molinolo, A.A., and Gutkind, J.S. (2015). Inactivation of a Gα(s)-PKA tumour suppressor pathway in skin stem cells initiates basal-cell carcinogenesis. Nat. Cell Biol. *17*, 793–803.

Jaffe, A.B., and Hall, A. (2005). Rho GTPases: biochemistry and biology. Annu. Rev. Cell Dev. Biol. *21*, 247–269.

Jia, J., Zhang, W., Wang, B., Trinko, R., and Jiang, J. (2003). The Drosophila Ste20 family kinase dMST functions as a tumor suppressor by restricting cell proliferation and promoting apoptosis. Genes Dev. *17*, 2514–2519.

Jiao, S., Wang, H., Shi, Z., Dong, A., Zhang, W., Song, X., He, F., Wang, Y., Zhang, Z., Wang, W., et al. (2014). A peptide mimicking VGLL4 function acts as a YAP antagonist therapy against gastric cancer. Cancer Cell *25*, 166–180.

Jin, Y., Xu, J., Yin, M.X., Lu, Y., Hu, L., Li, P., Zhang, P., Yuan, Z., Ho, M.S., Ji, H., et al. (2013). Brahma is essential for Drosophila intestinal stem cell proliferation and regulated by Hippo signaling. eLife *2*, e00999.

Judson, R.N., Tremblay, A.M., Knopp, P., White, R.B., Urcia, R., De Bari, C., Zammit, P.S., Camargo, F.D., and Wackerhage, H. (2012). The Hippo pathway member Yap plays a key role in influencing fate decisions in muscle satellite cells. J. Cell Sci. *125*, 6009–6019.

Jukam, D., Xie, B., Rister, J., Terrell, D., Charlton-Perkins, M., Pistillo, D., Gebelein, B., Desplan, C., and Cook, T. (2013). Opposite feedbacks in the Hippo pathway for growth control and neural fate. Science *342*, 1238016.

Justice, R.W., Zilian, O., Woods, D.F., Noll, M., and Bryant, P.J. (1995). The Drosophila tumor suppressor gene warts encodes a homolog of human myotonic dystrophy kinase and is required for the control of cell shape and proliferation. Genes Dev. *9*, 534–546.

Kango-Singh, M., Nolo, R., Tao, C., Verstreken, P., Hiesinger, P.R., Bellen, H.J., and Halder, G. (2002). Shar-pei mediates cell proliferation arrest during imaginal disc growth in Drosophila. Development *129*, 5719–5730.

Kapoor, A., Yao, W., Ying, H., Hua, S., Liewen, A., Wang, Q., Zhong, Y., Wu, C.J., Sadanandam, A., Hu, B., et al. (2014). Yap1 activation enables bypass of oncogenic Kras addiction in pancreatic cancer. Cell *158*, 185–197.

Karpowicz, P., Perez, J., and Perrimon, N. (2010). The Hippo tumor suppressor pathway regulates intestinal stem cell regeneration. Development *137*, 4135–4145.

Kim, M., Kim, M., Lee, S., Kuninaka, S., Saya, H., Lee, H., Lee, S., and Lim, D.S. (2013). cAMP/PKA signalling reinforces the LATS-YAP pathway to fully suppress YAP in response to actin cytoskeletal changes. EMBO J. *32*, 1543–1555.

Kim, K.M., Choi, Y.J., Hwang, J.H., Kim, A.R., Cho, H.J., Hwang, E.S., Park, J.Y., Lee, S.H., and Hong, J.H. (2014). Shear stress induced by an interstitial level of slow flow increases the osteogenic differentiation of mesenchymal stem cells through TAZ activation. PLoS ONE *9*, e92427.

Kim, M., Kim, T., Johnson, R.L., and Lim, D.S. (2015). Transcriptional corepressor function of the hippo pathway transducers YAP and TAZ. Cell Rep. *11*, 270–282.

Kono, K., Tamashiro, D.A., and Alarcon, V.B. (2014). Inhibition of RHO-ROCK signaling enhances ICM and suppresses TE characteristics through activation of Hippo signaling in the mouse blastocyst. Dev. Biol. *394*, 142–155.

Koontz, L.M., Liu-Chittenden, Y., Yin, F., Zheng, Y., Yu, J., Huang, B., Chen, Q., Wu, S., and Pan, D. (2013). The Hippo effector Yorkie controls normal tissue growth by antagonizing scalloped-mediated default repression. Dev. Cell *25*, 388–401.

Kwon, Y., Vinayagam, A., Sun, X., Dephoure, N., Gygi, S.P., Hong, P., and Perrimon, N. (2013). The Hippo signaling pathway interactome. Science *342*, 737–740.

Lai, Z.C., Wei, X., Shimizu, T., Ramos, E., Rohrbaugh, M., Nikolaidis, N., Ho, L.L., and Li, Y. (2005). Control of cell proliferation and apoptosis by mob as tumor suppressor, mats. Cell *120*, 675–685.

Lai, D., Ho, K.C., Hao, Y., and Yang, X. (2011). Taxol resistance in breast cancer cells is mediated by the hippo pathway component TAZ and its downstream transcriptional targets Cyr61 and CTGF. Cancer Res. *71*, 2728–2738.

Lange, A.W., Sridharan, A., Xu, Y., Stripp, B.R., Perl, A.K., and Whitsett, J.A. (2015). Hippo/Yap signaling controls epithelial progenitor cell proliferation and differentiation in the embryonic and adult lung. J. Mol. Cell Biol. *7*, 35–47.

Lavado, A., He, Y., Paré, J., Neale, G., Olson, E.N., Giovannini, M., and Cao, X. (2013). Tumor suppressor Nf2 limits expansion of the neural progenitor pool by inhibiting Yap/Taz transcriptional coactivators. Development *140*, 3323–3334.

Lavado, A., Ware, M., Paré, J., and Cao, X. (2014). The tumor suppressor Nf2 regulates corpus callosum development by inhibiting the transcriptional coactivator Yap. Development *141*, 4182–4193.

Lee, J.H., Kim, T.S., Yang, T.H., Koo, B.K., Oh, S.P., Lee, K.P., Oh, H.J., Lee, S.H., Kong, Y.Y., Kim, J.M., and Lim, D.S. (2008). A crucial role of WW45 in developing epithelial tissues in the mouse. EMBO J. *27*, 1231–1242.

Lee, K.P., Lee, J.H., Kim, T.S., Kim, T.H., Park, H.D., Byun, J.S., Kim, M.C., Jeong, W.I., Calvisi, D.F., Kim, J.M., and Lim, D.S. (2010). The Hippo-Salvador pathway restrains hepatic oval cell proliferation, liver size, and liver tumorigenesis. Proc. Natl. Acad. Sci. USA *107*, 8248–8253.

Lei, Q.Y., Zhang, H., Zhao, B., Zha, Z.Y., Bai, F., Pei, X.H., Zhao, S., Xiong, Y., and Guan, K.L. (2008). TAZ promotes cell proliferation and epithelial-mesenchymal transition and is inhibited by the hippo pathway. Mol. Cell. Biol. *28*, 2426–2436.

Li, S., Shen, D., Shao, J., Crowder, R., Liu, W., Prat, A., He, X., Liu, S., Hoog, J., Lu, C., et al. (2013). Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. Cell Rep. *4*, 1116–1130.

Li, Q., Li, S., Mana-Capelli, S., Roth Flach, R.J., Danai, L.V., Amcheslavsky, A., Nie, Y., Kaneko, S., Yao, X., Chen, X., et al. (2014). The conserved misshapen-warts-Yorkie pathway acts in enteroblasts to regulate intestinal stem cells in Drosophila. Dev. Cell *31*, 291–304.

Lian, I., Kim, J., Okazawa, H., Zhao, J., Zhao, B., Yu, J., Chinnaiyan, A., Israel, M.A., Goldstein, L.S., Abujarour, R., et al. (2010). The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. Genes Dev. *24*, 1106–1118.

Liang, N., Zhang, C., Dill, P., Panasyuk, G., Pion, D., Koka, V., Gallazzini, M., Olson, E.N., Lam, H., Henske, E.P., et al. (2014). Regulation of YAP by mTOR and autophagy reveals a therapeutic target of tuberous sclerosis complex. J. Exp. Med. *211*, 2249–2263.

Lin, Z., von Gise, A., Zhou, P., Gu, F., Ma, Q., Jiang, J., Yau, A.L., Buck, J.N., Gouin, K.A., van Gorp, P.R., et al. (2014). Cardiac-specific YAP activation improves cardiac function and survival in an experimental murine MI model. Circ. Res. *115*, 354–363.

Lin, C., Yao, E., and Chuang, P.T. (2015a). A conserved MST1/2-YAP axis mediates Hippo signaling during lung growth. Dev. Biol. *403*, 101–113.

Lin, L., Sabnis, A.J., Chan, E., Olivas, V., Cade, L., Pazarentzos, E., Asthana, S., Neel, D., Yan, J.J., Lu, X., et al. (2015b). The Hippo effector YAP promotes resistance to RAF- and MEK-targeted cancer therapies. Nat. Genet. *47*, 250–256.

Ling, C., Zheng, Y., Yin, F., Yu, J., Huang, J., Hong, Y., Wu, S., and Pan, D. (2010). The apical transmembrane protein Crumbs functions as a tumor suppressor that regulates Hippo signaling by binding to Expanded. Proc. Natl. Acad. Sci. USA *107*, 10532–10537.

Liu, C.Y., Zha, Z.Y., Zhou, X., Zhang, H., Huang, W., Zhao, D., Li, T., Chan, S.W., Lim, C.J., Hong, W., et al. (2010a). The hippo tumor pathway promotes TAZ degradation by phosphorylating a phosphodegron and recruiting the SCFbeta-TrCP E3 ligase. J. Biol. Chem. *285*, 37159–37169.

Liu, Y., Xin, Y., Ye, F., Wang, W., Lu, Q., Kaplan, H.J., and Dean, D.C. (2010b). Taz-tead1 links cell-cell contact to zeb1 expression, proliferation, and dedifferentiation in retinal pigment epithelial cells. Invest. Ophthalmol. Vis. Sci. *51*, 3372–3378.

Liu, G., Yu, F.X., Kim, Y.C., Meng, Z., Naipauer, J., Looney, D.J., Liu, X., Gutkind, J.S., Mesri, E.A., and Guan, K.L. (2015). Kaposi sarcoma-associated herpesvirus promotes tumorigenesis by modulating the Hippo pathway. Oncogene *34*, 3536–3546.

Liu-Chittenden, Y., Huang, B., Shim, J.S., Chen, Q., Lee, S.J., Anders, R.A., Liu, J.O., and Pan, D. (2012). Genetic and pharmacological disruption of the TEAD-YAP complex suppresses the oncogenic activity of YAP. Genes Dev. *26*, 1300–1305.

Lorthongpanich, C., Messerschmidt, D.M., Chan, S.W., Hong, W., Knowles, B.B., and Solter, D. (2013). Temporal reduction of LATS kinases in the early preimplantation embryo prevents ICM lineage differentiation. Genes Dev. *27*, 1441–1446.

Lu, L., Li, Y., Kim, S.M., Bossuyt, W., Liu, P., Qiu, Q., Wang, Y., Halder, G., Finegold, M.J., Lee, J.S., and Johnson, R.L. (2010). Hippo signaling is a potent in vivo growth and tumor suppressor pathway in the mammalian liver. Proc. Natl. Acad. Sci. USA *107*, 1437–1442.

Ma, B., Chen, Y., Chen, L., Cheng, H., Mu, C., Li, J., Gao, R., Zhou, C., Cao, L., Liu, J., et al. (2015). Hypoxia regulates Hippo signalling through the SIAH2 ubiquitin E3 ligase. Nat. Cell Biol. *17*, 95–103.

Mahoney, J.E., Mori, M., Szymaniak, A.D., Varelas, X., and Cardoso, W.V. (2014). The hippo pathway effector Yap controls patterning and differentiation of airway epithelial progenitors. Dev. Cell *30*, 137–150.

Makita, R., Uchijima, Y., Nishiyama, K., Amano, T., Chen, Q., Takeuchi, T., Mitani, A., Nagase, T., Yatomi, Y., Aburatani, H., et al. (2008). Multiple renal cysts, urinary concentration defects, and pulmonary emphysematous changes in mice lacking TAZ. Am. J. Physiol. Renal Physiol. *294*, F542–F553.

Matakatsu, H., and Blair, S.S. (2012). Separating planar cell polarity and Hippo pathway activities of the protocadherins Fat and Dachsous. Development *139*, 1498–1508.

McClatchey, A.I., Saotome, I., Mercer, K., Crowley, D., Gusella, J.F., Bronson, R.T., and Jacks, T. (1998). Mice heterozygous for a mutation at the Nf2 tumor suppressor locus develop a range of highly metastatic tumors. Genes Dev. *12*, 1121–1133.

Meng, Z., Moroishi, T., Mottier-Pavie, V., Plouffe, S.W., Hansen, C.G., Hong, A.W., Park, H.W., Mo, J.S., Lu, W., Lu, S., et al. (2015). MAP4K family kinases act in parallel to MST1/2 to activate LATS1/2 in the Hippo pathway. Nat. Commun. *6*, 8357.

Mi, W., Lin, Q., Childress, C., Sudol, M., Robishaw, J., Berlot, C.H., Shabahang, M., and Yang, W. (2015). Geranylgeranylation signals to the Hippo pathway for breast cancer cell proliferation and migration. Oncogene *34*, 3095–3106.

Miller, E., Yang, J., DeRan, M., Wu, C., Su, A.I., Bonamy, G.M., Liu, J., Peters, E.C., and Wu, X. (2012). Identification of serum-derived sphingosine-1-phosphate as a small molecule regulator of YAP. Chem. Biol. *19*, 955–962.

Mitani, A., Nagase, T., Fukuchi, K., Aburatani, H., Makita, R., and Kurihara, H. (2009). Transcriptional coactivator with PDZ-binding motif is essential for normal alveolarization in mice. Am. J. Respir. Crit. Care Med. *180*, 326–338.

Miyanaga, A., Masuda, M., Tsuta, K., Kawasaki, K., Nakamura, Y., Sakuma, T., Asamura, H., Gemma, A., and Yamada, T. (2015). Hippo pathway gene mutations in malignant mesothelioma: revealed by RNA and targeted exon sequencing. J. Thorac. Oncol. *10*, 844–851.

Mo, J.S., Yu, F.X., Gong, R., Brown, J.H., and Guan, K.L. (2012). Regulation of the Hippo-YAP pathway by protease-activated receptors (PARs). Genes Dev. *26*, 2138–2143.

Mo, J.S., Meng, Z., Kim, Y.C., Park, H.W., Hansen, C.G., Kim, S., Lim, D.S., and Guan, K.L. (2015). Cellular energy stress induces AMPK-mediated regulation of YAP and the Hippo pathway. Nat. Cell Biol. *17*, 500–510.

Mohseni, M., Sun, J., Lau, A., Curtis, S., Goldsmith, J., Fox, V.L., Wei, C., Frazier, M., Samson, O., Wong, K.K., et al. (2014). A genetic screen identifies an LKB1-MARK signalling axis controlling the Hippo-YAP pathway. Nat. Cell Biol. *16*, 108–117.

Mori, M., Triboulet, R., Mohseni, M., Schlegelmilch, K., Shrestha, K., Camargo, F.D., and Gregory, R.I. (2014). Hippo signaling regulates microprocessor and links cell-density-dependent miRNA biogenesis to cancer. Cell *156*, 893–906.

Morin-Kensicki, E.M., Boone, B.N., Howell, M., Stonebraker, J.R., Teed, J., Alb, J.G., Magnuson, T.R., O'Neal, W., and Milgram, S.L. (2006). Defects in yolk sac vasculogenesis, chorioallantoic fusion, and embryonic axis elongation in mice with targeted disruption of Yap65. Mol. Cell. Biol. *26*, 77–87.

Moroishi, T., Hansen, C.G., and Guan, K.L. (2015a). The emerging roles of YAP and TAZ in cancer. Nat. Rev. Cancer *15*, 73–79.

Moroishi, T., Park, H.W., Qin, B., Chen, Q., Meng, Z., Plouffe, S.W., Taniguchi, K., Yu, F.X., Karin, M., Pan, D., and Guan, K.L. (2015b). A YAP/TAZ-induced feedback mechanism regulates Hippo pathway homeostasis. Genes Dev. 29, 1271–1284.

Murakami, H., Mizuno, T., Taniguchi, T., Fujii, M., Ishiguro, F., Fukui, T., Akatsuka, S., Horio, Y., Hida, T., Kondo, Y., et al. (2011). LATS2 is a tumor suppressor gene of malignant mesothelioma. Cancer Res. 71, 873–883.

Nishio, M., Hamada, K., Kawahara, K., Sasaki, M., Noguchi, F., Chiba, S., Mizuno, K., Suzuki, S.O., Dong, Y., Tokuda, M., et al. (2012). Cancer susceptibility and embryonic lethality in Mob1a/1b double-mutant mice. J. Clin. Invest. 122, 4505–4518.

Nishioka, N., Inoue, K., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R.O., Ogonuki, N., et al. (2009). The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. Dev. Cell 16, 398–410.

Nolo, R., Morrison, C.M., Tao, C., Zhang, X., and Halder, G. (2006). The bantam microRNA is a target of the hippo tumor-suppressor pathway. Curr. Biol. 16, 1895–1904.

Oh, H., and Irvine, K.D. (2008). In vivo regulation of Yorkie phosphorylation and localization. Development 135, 1081–1088.

Oh, H., Slattery, M., Ma, L., Crofts, A., White, K.P., Mann, R.S., and Irvine, K.D. (2013). Genome-wide association of Yorkie with chromatin and chromatin-remodeling complexes. Cell Rep. 3, 309–318.

Ota, M., and Sasaki, H. (2008). Mammalian Tead proteins regulate cell proliferation and contact inhibition as transcriptional mediators of Hippo signaling. Development 135, 4059–4069.

Overholtzer, M., Zhang, J., Smolen, G.A., Muir, B., Li, W., Sgroi, D.C., Deng, C.X., Brugge, J.S., and Haber, D.A. (2006). Transforming properties of YAP, a candidate oncogene on the chromosome 11q22 amplicon. Proc. Natl. Acad. Sci. USA 103, 12405–12410.

Pajtler, K.W., Witt, H., Sill, M., Jones, D.T.W., Hovestadt, V., Kratochwil, F., Wani, K., Tatevossian, R., Punchihewa, C., Johann, P., et al. (2015). Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. Cancer Cell 27, 728–743.

Pan, D. (2010). The hippo signaling pathway in development and cancer. Dev. Cell 19, 491–505.

Pantalacci, S., Tapon, N., and Léopold, P. (2003). The Salvador partner Hippo promotes apoptosis and cell-cycle exit in Drosophila. Nat. Cell Biol. 5, 921–927.

Park, H.W., and Guan, K.L. (2013). Regulation of the Hippo pathway and implications for anticancer drug development. Trends Pharmacol. Sci. 34, 581–589.

Park, H.W., Kim, Y.C., Yu, B., Moroishi, T., Mo, J.S., Plouffe, S.W., Meng, Z., Lin, K.C., Yu, F.X., Alexander, C.M., et al. (2015). Alternative Wnt Signaling Activates YAP/TAZ. Cell 162, 780–794.

Parker, M., Mohankumar, K.M., Punchihewa, C., Weinlich, R., Dalton, J.D., Li, Y., Lee, R., Tatevossian, R.G., Phoenix, T.N., Thiruvenkatam, R., et al. (2014). C11orf95-RELA fusions drive oncogenic NF-κB signalling in ependymoma. Nature 506, 451–455.

Poon, C.L., Lin, J.I., Zhang, X., and Harvey, K.F. (2011). The sterile 20-like kinase Tao-1 controls tissue growth by regulating the Salvador-Warts-Hippo pathway. Dev. Cell 21, 896–906.

Porazinski, S., Wang, H., Asaoka, Y., Behrndt, M., Miyamoto, T., Morita, H., Hata, S., Sasaki, T., Krens, S.F., Osada, Y., et al. (2015). YAP is essential for tissue tension to ensure vertebrate 3D body shape. Nature 521, 217–221.

Praskova, M., Khoklatchev, A., Ortiz-Vega, S., and Avruch, J. (2004). Regulation of the MST1 kinase by autophosphorylation, by the growth inhibitory proteins, RASSF1 and NORE1, and by Ras. Biochem. J. 381, 453–462.

Qing, Y., Yin, F., Wang, W., Zheng, Y., Guo, P., Schozer, F., Deng, H., and Pan, D. (2014). The Hippo effector Yorkie activates transcription by interacting with a histone methyltransferase complex through Ncoa6. eLife 3, e02564.

Rauskolb, C., Sun, S., Sun, G., Pan, Y., and Irvine, K.D. (2014). Cytoskeletal tension inhibits Hippo signaling through an Ajuba-Warts complex. Cell 158, 143–156.

Reddy, B.V., and Irvine, K.D. (2013). Regulation of Hippo signaling by EGFR-MAPK signaling through Ajuba family proteins. Dev. Cell 24, 459–471.

Reginensi, A., Scott, R.P., Gregorieff, A., Bagherie-Lachidan, M., Chung, C., Lim, D.S., Pawson, T., Wrana, J., and McNeill, H. (2013). Yap- and Cdc42-dependent nephrogenesis and morphogenesis during mouse kidney development. PLoS Genet. 9, e1003380.

Ren, F., Wang, B., Yue, T., Yun, E.Y., Ip, Y.T., and Jiang, J. (2010). Hippo signaling regulates Drosophila intestine stem cell proliferation through multiple pathways. Proc. Natl. Acad. Sci. USA 107, 21064–21069.

Robinson, B.S., Huang, J., Hong, Y., and Moberg, K.H. (2010). Crumbs regulates Salvador/Warts/Hippo signaling in Drosophila via the FERM-domain protein Expanded. Curr. Biol. 20, 582–590.

Rouleau, G.A., Merel, P., Lutchman, M., Sanson, M., Zucman, J., Marineau, C., Hoang-Xuan, K., Demczuk, S., Desmaze, C., Plougastel, B., et al. (1993). Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. Nature 363, 515–521.

Sansores-Garcia, L., Bossuyt, W., Wada, K., Yonemura, S., Tao, C., Sasaki, H., and Halder, G. (2011). Modulating F-actin organization induces organ growth by affecting the Hippo pathway. EMBO J. 30, 2325–2335.

Sasaki, H. (2015). Position- and polarity-dependent Hippo signaling regulates cell fates in preimplantation mouse embryos. Semin. Cell Dev. Biol. S1084-9521(15)00100-7.

Schlegelmilch, K., Mohseni, M., Kirak, O., Pruszak, J., Rodriguez, J.R., Zhou, D., Kreger, B.T., Vasioukhin, V., Avruch, J., Brummelkamp, T.R., and Camargo, F.D. (2011). Yap1 acts downstream of α-catenin to control epidermal proliferation. Cell 144, 782–795.

Schrauwen, I., Szelinger, S., Siniard, A.L., Corneveaux, J.J., Kurdoglu, A., Richholt, R., De Both, M., Malenica, I., Swaminathan, S., Rangasamy, S., et al. (2015). A De Novo Mutation in TEAD1 Causes Non-X-Linked Aicardi Syndrome. Invest. Ophthalmol. Vis. Sci. 56, 3896–3904.

Sekido, Y. (2011). Inactivation of Merlin in malignant mesothelioma cells and the Hippo signaling cascade dysregulation. Pathol. Int. 61, 331–344.

Shao, D.D., Xue, W., Krall, E.B., Bhutkar, A., Piccioni, F., Wang, X., Schinzel, A.C., Sood, S., Rosenbluh, J., Kim, J.W., et al. (2014). KRAS and YAP1 converge to regulate EMT and tumor survival. Cell 158, 171–184.

Sharma, P., and McNeill, H. (2013). Fat and Dachsous cadherins. Prog. Mol. Biol. Transl. Sci. 116, 215–235.

Shaw, R.L., Kohlmaier, A., Polesello, C., Veelken, C., Edgar, B.A., and Tapon, N. (2010). The Hippo pathway regulates intestinal stem cell proliferation during Drosophila adult midgut regeneration. Development 137, 4147–4158.

Shen, S., Guo, X., Yan, H., Lu, Y., Ji, X., Li, L., Liang, T., Zhou, D., Feng, X.H., Zhao, J.C., et al. (2015). A miR-130a-YAP positive feedback loop promotes organ size and tumorigenesis. Cell Res. 25, 997–1012.

Silva, E., Tsatskis, Y., Gardano, L., Tapon, N., and McNeill, H. (2006). The tumor-suppressor gene fat controls tissue growth upstream of expanded in the hippo signaling pathway. Curr. Biol. 16, 2081–2089.

Skibinski, A., Breindel, J.L., Prat, A., Galván, P., Smith, E., Rolfs, A., Gupta, P.B., Labaer, J., and Kuperwasser, C. (2014). The Hippo transducer TAZ interacts with the SWI/SNF complex to regulate breast epithelial lineage commitment. Cell Rep. 6, 1059–1072.

Song, H., Mak, K.K., Topol, L., Yun, K., Hu, J., Garrett, L., Chen, Y., Park, O., Chang, J., Simpson, R.M., et al. (2010). Mammalian Mst1 and Mst2 kinases play essential roles in organ size control and tumor suppression. Proc. Natl. Acad. Sci. USA 107, 1431–1436.

Song, Y., Li, L., Ou, Y., Gao, Z., Li, E., Li, X., Zhang, W., Wang, J., Xu, L., Zhou, Y., et al. (2014). Identification of genomic alterations in oesophageal squamous cell cancer. Nature 509, 91–95.

Sorrentino, G., Ruggeri, N., Specchia, V., Cordenonsi, M., Mano, M., Dupont, S., Manfrin, A., Ingallina, E., Sommaggio, R., Piazza, S., et al. (2014). Metabolic control of YAP and TAZ by the mevalonate pathway. Nat. Cell Biol. 16, 357–366.

Stein, C., Bardet, A.F., Roma, G., Bergling, S., Clay, I., Ruchti, A., Agarinis, C., Schmelzle, T., Bouwmeester, T., Schübeler, D., and Bauer, A. (2015). YAP1

Exerts Its Transcriptional Control via TEAD-Mediated Activation of Enhancers. PLoS Genet. *11*, e1005465.

Steinhardt, A.A., Gayyed, M.F., Klein, A.P., Dong, J., Maitra, A., Pan, D., Montgomery, E.A., and Anders, R.A. (2008). Expression of Yes-associated protein in common solid tumors. Hum. Pathol. *39*, 1582–1589.

Straßburger, K., Tiebe, M., Pinna, F., Breuhahn, K., and Teleman, A.A. (2012). Insulin/IGF signaling drives cell proliferation in part via Yorkie/YAP. Dev. Biol. *367*, 187–196.

Su, T., Bondar, T., Zhou, X., Zhang, C., He, H., and Medzhitov, R. (2015). Two-signal requirement for growth-promoting function of Yap in hepatocytes. eLife *4*, e02948.

Tanas, M.R., Sboner, A., Oliveira, A.M., Erickson-Johnson, M.R., Hespelt, J., Hanwright, P.J., Flanagan, J., Luo, Y., Fenwick, K., Natrajan, R., et al. (2011). Identification of a disease-defining gene fusion in epithelioid hemangioendothelioma. Sci. Transl. Med. *3*, 98ra82.

Tapon, N., Harvey, K.F., Bell, D.W., Wahrer, D.C., Schiripo, T.A., Haber, D., and Hariharan, I.K. (2002). salvador Promotes both cell cycle exit and apoptosis in Drosophila and is mutated in human cancer cell lines. Cell *110*, 467–478.

Thompson, B.J., and Cohen, S.M. (2006). The Hippo pathway regulates the bantam microRNA to control cell proliferation and apoptosis in Drosophila. Cell *126*, 767–774.

Tian, Y., Kolb, R., Hong, J.H., Carroll, J., Li, D., You, J., Bronson, R., Yaffe, M.B., Zhou, J., and Benjamin, T. (2007). TAZ promotes PC2 degradation through a SCFbeta-Trcp E3 ligase complex. Mol. Cell. Biol. *27*, 6383–6395.

Touil, Y., Igoudjil, W., Corvaisier, M., Dessein, A.F., Vandomme, J., Monté, D., Stechly, L., Skrypek, N., Langlois, C., Grard, G., et al. (2014). Colon cancer cells escape 5FU chemotherapy-induced cell death by entering stemness and quiescence associated with the c-Yes/YAP axis. Clin. Cancer Res. *20*, 837–846.

Tyler, D.M., and Baker, N.E. (2007). Expanded and fat regulate growth and differentiation in the Drosophila eye through multiple signaling pathways. Dev. Biol. *305*, 187–201.

Udan, R.S., Kango-Singh, M., Nolo, R., Tao, C., and Halder, G. (2003). Hippo promotes proliferation arrest and apoptosis in the Salvador/Warts pathway. Nat. Cell Biol. *5*, 914–920.

Van Raamsdonk, C.D., Bezrookove, V., Green, G., Bauer, J., Gaugler, L., O'Brien, J.M., Simpson, E.M., Barsh, G.S., and Bastian, B.C. (2009). Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. Nature *457*, 599–602.

Van Raamsdonk, C.D., Griewank, K.G., Crosby, M.B., Garrido, M.C., Vemula, S., Wiesner, T., Obenauf, A.C., Wackernagel, W., Green, G., Bouvier, N., et al. (2010). Mutations in GNA11 in uveal melanoma. N. Engl. J. Med. *363*, 2191–2199.

Varelas, X. (2014). The Hippo pathway effectors TAZ and YAP in development, homeostasis and disease. Development *141*, 1614–1626.

von Gise, A., Lin, Z., Schlegelmilch, K., Honor, L.B., Pan, G.M., Buck, J.N., Ma, Q., Ishiwata, T., Zhou, B., Camargo, F.D., and Pu, W.T. (2012). YAP1, the nuclear target of Hippo signaling, stimulates heart growth through cardiomyocyte proliferation but not hypertrophy. Proc. Natl. Acad. Sci. USA *109*, 2394–2399.

Wada, K., Itoga, K., Okano, T., Yonemura, S., and Sasaki, H. (2011). Hippo pathway regulation by cell morphology and stress fibers. Development *138*, 3907–3914.

Wang, Z., Wu, Y., Wang, H., Zhang, Y., Mei, L., Fang, X., Zhang, X., Zhang, F., Chen, H., Liu, Y., et al. (2014). Interplay of mevalonate and Hippo pathways regulates RHAMM transcription via YAP to modulate breast cancer cell motility. Proc. Natl. Acad. Sci. USA *111*, E89–E98.

Wang, W., Xiao, Z.D., Li, X., Aziz, K.E., Gan, B., Johnson, R.L., and Chen, J. (2015). AMPK modulates Hippo pathway activity to regulate energy homeostasis. Nat. Cell Biol. *17*, 490–499.

Watt, K.I., Turner, B.J., Hagg, A., Zhang, X., Davey, J.R., Qian, H., Beyer, C., Winbanks, C.E., Harvey, K.F., and Gregorevic, P. (2015). The Hippo pathway effector YAP is a critical regulator of skeletal muscle fibre size. Nat. Commun. *6*, 6048.

Wehr, M.C., Holder, M.V., Gailite, I., Saunders, R.E., Maile, T.M., Ciirdaeva, E., Instrell, R., Jiang, M., Howell, M., Rossner, M.J., and Tapon, N. (2013). Salt-inducible kinases regulate growth through the Hippo signalling pathway in Drosophila. Nat. Cell Biol. *15*, 61–71.

Wennmann, D.O., Vollenbroker, B., Eckart, A.K., Bonse, J., Erdmann, F., Wolters, D.A., Schenk, L.K., Schulze, U., Kremerskothen, J., Weide, T., et al. (2014). The Hippo pathway is controlled by Angiotensin II signaling and its reactivation induces apoptosis in podocytes. Cell Death Dis. *5*, e1519.

Willecke, M., Hamaratoglu, F., Kango-Singh, M., Udan, R., Chen, C.L., Tao, C., Zhang, X., and Halder, G. (2006). The fat cadherin acts through the hippo tumor-suppressor pathway to regulate tissue size. Curr. Biol. *16*, 2090–2100.

Williamson, K.A., Rainger, J., Floyd, J.A., Ansari, M., Meynert, A., Aldridge, K.V., Rainger, J.K., Anderson, C.A., Moore, A.T., Hurles, M.E., et al.; UK10K Consortium (2014). Heterozygous loss-of-function mutations in YAP1 cause both isolated and syndromic optic fissure closure defects. Am. J. Hum. Genet. *94*, 295–302.

Wong, K.K., Li, W., An, Y., Duan, Y., Li, Z., Kang, Y., and Yan, Y. (2015). β-Spectrin regulates the hippo signaling pathway and modulates the basal actin network. J. Biol. Chem. *290*, 6397–6407.

Wu, S., Huang, J., Dong, J., and Pan, D. (2003). hippo encodes a Ste-20 family protein kinase that restricts cell proliferation and promotes apoptosis in conjunction with salvador and warts. Cell *114*, 445–456.

Wu, S., Liu, Y., Zheng, Y., Dong, J., and Pan, D. (2008). The TEAD/TEF family protein Scalloped mediates transcriptional output of the Hippo growth-regulatory pathway. Dev. Cell *14*, 388–398.

Wu, H., Xiao, Y., Zhang, S., Ji, S., Wei, L., Fan, F., Geng, J., Tian, J., Sun, X., Qin, F., et al. (2013). The Ets transcription factor GABP is a component of the hippo pathway essential for growth and antioxidant defense. Cell Rep. *3*, 1663–1677.

Xiang, L., Gilkes, D.M., Hu, H., Takano, N., Luo, W., Lu, H., Bullen, J.W., Samanta, D., Liang, H., and Semenza, G.L. (2014). Hypoxia-inducible factor 1 mediates TAZ expression and nuclear localization to induce the breast cancer stem cell phenotype. Oncotarget *5*, 12509–12527.

Xiao, G.H., Chernoff, J., and Testa, J.R. (2003). NF2: the wizardry of merlin. Genes Chromosomes Cancer *38*, 389–399.

Xin, M., Kim, Y., Sutherland, L.B., Qi, X., McAnally, J., Schwartz, R.J., Richardson, J.A., Bassel-Duby, R., and Olson, E.N. (2011). Regulation of insulin-like growth factor signaling by Yap governs cardiomyocyte proliferation and embryonic heart size. Sci. Signal. *4*, ra70.

Xin, M., Kim, Y., Sutherland, L.B., Murakami, M., Qi, X., McAnally, J., Porrello, E.R., Mahmoud, A.I., Tan, W., Shelton, J.M., et al. (2013). Hippo pathway effector Yap promotes cardiac regeneration. Proc. Natl. Acad. Sci. USA *110*, 13839–13844.

Xu, T., Wang, W., Zhang, S., Stewart, R.A., and Yu, W. (1995). Identifying tumor suppressors in genetic mosaics: the Drosophila lats gene encodes a putative protein kinase. Development *121*, 1053–1063.

Yimlamai, D., Christodoulou, C., Galli, G.G., Yanger, K., Pepe-Mooney, B., Gurung, B., Shrestha, K., Cahan, P., Stanger, B.Z., and Camargo, F.D. (2014). Hippo pathway activity influences liver cell fate. Cell *157*, 1324–1338.

Yin, F., Yu, J., Zheng, Y., Chen, Q., Zhang, N., and Pan, D. (2013). Spatial organization of Hippo signaling at the plasma membrane mediated by the tumor suppressor Merlin/NF2. Cell *154*, 1342–1355.

Yu, F.X., and Guan, K.L. (2013). The Hippo pathway: regulators and regulations. Genes Dev. *27*, 355–371.

Yu, J., Zheng, Y., Dong, J., Klusza, S., Deng, W.M., and Pan, D. (2010). Kibra functions as a tumor suppressor protein that regulates Hippo signaling in conjunction with Merlin and Expanded. Dev. Cell *18*, 288–299.

Yu, F.X., Zhao, B., Panupinthu, N., Jewell, J.L., Lian, I., Wang, L.H., Zhao, J., Yuan, H., Tumaneng, K., Li, H., et al. (2012). Regulation of the Hippo-YAP pathway by G-protein-coupled receptor signaling. Cell *150*, 780–791.

Yu, F.X., Zhang, Y., Park, H.W., Jewell, J.L., Chen, Q., Deng, Y., Pan, D., Taylor, S.S., Lai, Z.C., and Guan, K.L. (2013a). Protein kinase A activates the Hippo pathway to modulate cell proliferation and differentiation. Genes Dev. *27*, 1223–1232.

Yu, T., Bachman, J., and Lai, Z.C. (2013b). Evidence for a tumor suppressor role for the large tumor suppressor genes LATS1 and LATS2 in human cancer. Genetics *195*, 1193–1196.

Yu, F.X., Luo, J., Mo, J.S., Liu, G., Kim, Y.C., Meng, Z., Zhao, L., Peyman, G., Ouyang, H., Jiang, W., et al. (2014). Mutant Gq/11 promote uveal melanoma tumorigenesis by activating YAP. Cancer Cell *25*, 822–830.

Yu, F.X., Meng, Z., Plouffe, S.W., and Guan, K.L. (2015). Hippo pathway regulation of gastrointestinal tissues. Annu. Rev. Physiol. *77*, 201–227.

Yue, T., Tian, A., and Jiang, J. (2012). The cell adhesion molecule echinoid functions as a tumor suppressor and upstream regulator of the Hippo signaling pathway. Dev. Cell *22*, 255–267.

Zanconato, F., Forcato, M., Battilana, G., Azzolin, L., Quaranta, E., Bodega, B., Rosato, A., Bicciato, S., Cordenonsi, M., and Piccolo, S. (2015). Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. Nat. Cell Biol. *17*, 1218–1227.

Zender, L., Spector, M.S., Xue, W., Flemming, P., Cordon-Cardo, C., Silke, J., Fan, S.T., Luk, J.M., Wigler, M., Hannon, G.J., et al. (2006). Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. Cell *125*, 1253–1267.

Zhang, L., Ren, F., Zhang, Q., Chen, Y., Wang, B., and Jiang, J. (2008). The TEAD/TEF family of transcription factor Scalloped mediates Hippo signaling in organ size control. Dev. Cell *14*, 377–387.

Zhang, N., Bai, H., David, K.K., Dong, J., Zheng, Y., Cai, J., Giovannini, M., Liu, P., Anders, R.A., and Pan, D. (2010). The Merlin/NF2 tumor suppressor functions through the YAP oncoprotein to regulate tissue homeostasis in mammals. Dev. Cell *19*, 27–38.

Zhang, H., Pasolli, H.A., and Fuchs, E. (2011). Yes-associated protein (YAP) transcriptional coactivator functions in balancing growth and differentiation in skin. Proc. Natl. Acad. Sci. USA *108*, 2270–2275.

Zhang, W., Nandakumar, N., Shi, Y., Manzano, M., Smith, A., Graham, G., Gupta, S., Vietsch, E.E., Laughlin, S.Z., Wadhwa, M., et al. (2014a). Downstream of mutant KRAS, the transcription regulator YAP is essential for neoplastic progression to pancreatic ductal adenocarcinoma. Sci. Signal. *7*, ra42.

Zhang, W., Gao, Y., Li, P., Shi, Z., Guo, T., Li, F., Han, X., Feng, Y., Zheng, C., Wang, Z., et al. (2014b). VGLL4 functions as a new tumor suppressor in lung cancer by negatively regulating the YAP-TEAD transcriptional complex. Cell Res. *24*, 331–343.

Zhao, B., Wei, X., Li, W., Udan, R.S., Yang, Q., Kim, J., Xie, J., Ikenoue, T., Yu, J., Li, L., et al. (2007). Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. Genes Dev. *21*, 2747–2761.

Zhao, B., Ye, X., Yu, J., Li, L., Li, W., Li, S., Yu, J., Lin, J.D., Wang, C.Y., Chinnaiyan, A.M., et al. (2008). TEAD mediates YAP-dependent gene induction and growth control. Genes Dev. *22*, 1962–1971.

Zhao, B., Li, L., Tumaneng, K., Wang, C.Y., and Guan, K.L. (2010). A coordinated phosphorylation by Lats and CK1 regulates YAP stability through SCF(beta-TRCP). Genes Dev. *24*, 72–85.

Zhao, B., Li, L., Wang, L., Wang, C.Y., Yu, J., and Guan, K.L. (2012). Cell detachment activates the Hippo pathway via cytoskeleton reorganization to induce anoikis. Genes Dev. *26*, 54–68.

Zheng, Y., Wang, W., Liu, B., Deng, H., Uster, E., and Pan, D. (2015). Identification of Happyhour/MAP4K as Alternative Hpo/Mst-like Kinases in the Hippo Kinase Cascade. Dev. Cell *34*, 642–655.

Zhou, D., Conrad, C., Xia, F., Park, J.S., Payer, B., Yin, Y., Lauwers, G.Y., Thasler, W., Lee, J.T., Avruch, J., and Bardeesy, N. (2009). Mst1 and Mst2 maintain hepatocyte quiescence and suppress hepatocellular carcinoma development through inactivation of the Yap1 oncogene. Cancer Cell *16*, 425–438.

Zhou, D., Zhang, Y., Wu, H., Barry, E., Yin, Y., Lawrence, E., Dawson, D., Willis, J.E., Markowitz, S.D., Camargo, F.D., and Avruch, J. (2011). Mst1 and Mst2 protein kinases restrain intestinal stem cell proliferation and colonic tumorigenesis by inhibition of Yes-associated protein (Yap) overabundance. Proc. Natl. Acad. Sci. USA *108*, E1312–E1320.

Zhou, X., Wang, S., Wang, Z., Feng, X., Liu, P., Lv, X.B., Li, F., Yu, F.X., Sun, Y., Yuan, H., et al. (2015). Estrogen regulates Hippo signaling via GPER in breast cancer. J. Clin. Invest. *125*, 2123–2135.

The LC Domain of hnRNPA2 Adopts Similar Conformations in Hydrogel Polymers, Liquid-like Droplets, and Nuclei

Graphical Abstract
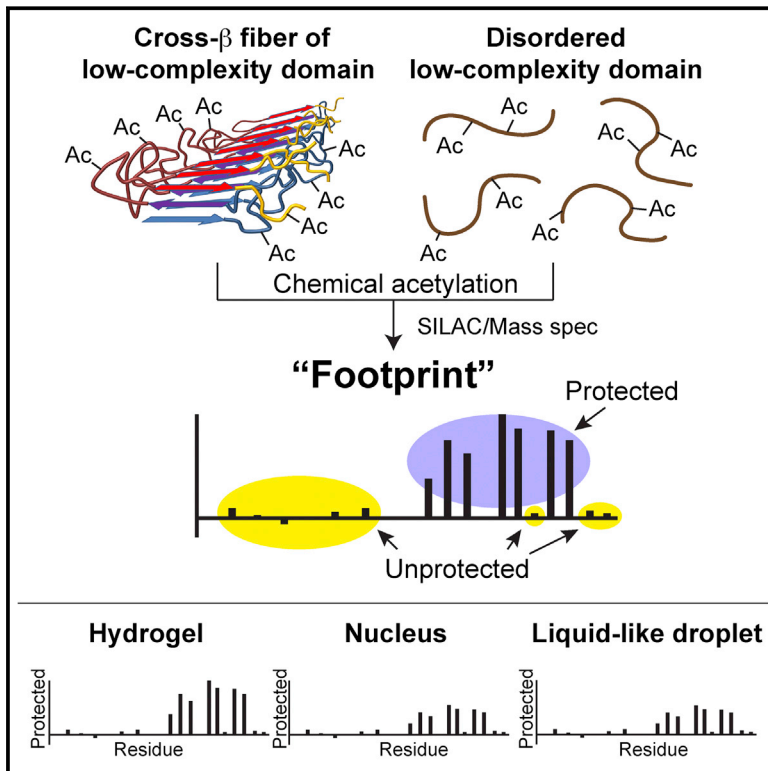
Authors

Siheng Xiang, Masato Kato, Leeju C. Wu, ..., Yajie Zhang, Yonghao Yu, Steven L. McKnight

Correspondence

yonghao.yu@utsouthwestern.edu (Y.Y.), steven.mcknight@utsouthwestern.edu (S.L.M.)

In Brief

A chemical footprinting method reveals that polymers of low-complexity domains exhibit similar cross-β structure in hydrogels, liquid-like droplets, and nuclei of mammalian cells, suggesting a common underlying structural basis.

Highlights

- A footprinting method was used to probe cross-β structure of LC domain polymers

- Similar footprints were obtained from hydrogels, liquid-like droplets, and nuclei

- Mutations impeding hydrogel binding map to the core of the LC domain footprint

- Hydrogel and liquid-like droplet formation is driven by cross-β polymerization

CellPress

Article

The LC Domain of hnRNPA2 Adopts Similar Conformations in Hydrogel Polymers, Liquid-like Droplets, and Nuclei

Siheng Xiang,[1] Masato Kato,[1] Leeju C. Wu,[1] Yi Lin,[1] Ming Ding,[1] Yajie Zhang,[1] Yonghao Yu,[1,*] and Steven L. McKnight[1,*]
[1]Department of Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390, USA
*Correspondence: yonghao.yu@utsouthwestern.edu (Y.Y.), steven.mcknight@utsouthwestern.edu (S.L.M.)
http://dx.doi.org/10.1016/j.cell.2015.10.040

SUMMARY

Many DNA and RNA regulatory proteins contain polypeptide domains that are unstructured when analyzed in cell lysates. These domains are typified by an over-representation of a limited number of amino acids and have been termed prion-like, intrinsically disordered or low-complexity (LC) domains. When incubated at high concentration, certain of these LC domains polymerize into labile, amyloid-like fibers. Here, we report methods allowing the generation of a molecular footprint of the polymeric state of the LC domain of hnRNPA2. By deploying this footprinting technique to probe the structure of the native hnRNPA2 protein present in isolated nuclei, we offer evidence that its LC domain exists in a similar conformation as that described for recombinant polymers of the protein. These observations favor biologic utility to the polymerization of LC domains in the pathway of information transfer from gene to message to protein.

INTRODUCTION

DNA and RNA regulatory proteins are composed of two functional domains. Gene-specific transcription factors contain DNA binding domains that recognize specific sequences via structurally ordered states, including zinc fingers, homeoboxes, helix-loop-helix domains, and leucine zipper domains (Pabo and Sauer, 1992). Likewise, RNA binding proteins are able to bind RNA via structurally ordered KH domains, RNA recognition motifs, and pumilio domains (Lunde et al., 2007).

Most DNA and RNA regulatory proteins also contain polypeptide domains that lack structural order when purified from cellular lysates. The unstructured activation domains of certain transcription factors contain an over-representation of acidic amino acids (Hope et al., 1988). In the context of gene-specific transcription factors, these structurally disordered domains have been termed "acid blobs" or "negative noodles" (Sigler, 1988) and other conceptualizations invoking biological function in the absence of folded protein structure.

Not all activation domains associated with gene specific transcription factors are acidic. Some are enriched in glutamine residues and others in proline residues (Triezenberg, 1995).

Common, however, among the majority of activation domains is the over-representation of one or a small grouping of amino acids. Instead of utilizing a balanced proportion of all 20 amino acids, these domains are of low complexity in nature. Nucleic acids deploy a four-lettered code, proteins a 20 letter code. Low-complexity (LC) domains operate via the deployment of a highly skewed distribution of amino acids and would appear to be much more DNA and RNA like in the nature of their code.

RNA binding proteins also contain LC domains, including repetitive polymers of serine and arginine (SR) in many proteins that regulate pre-mRNA splicing (Manley and Tacke, 1996), and G/S-Y-G/S repeats in the LC domains associated with the FET, CIRBP/RBM3, and hnRNP families of RNA binding proteins (Kato et al., 2012). Compared with gene-specific transcription factors and their activation domains, less attention has been paid to the LC domains associated with RNA regulatory proteins. Some degree of attention has been focused on the LC domains associated with the FET family of RNA binding proteins, including fused in sarcoma (FUS), Ewing's sarcoma (EWS), and TAF15. The amino terminal LC domains of these three proteins can be translocated onto DNA binding domains as the causative event in many forms of human cancer (Riggi et al., 2007). In the context of these fusion proteins, the LC domains of the FET proteins are understood to function as potent transcriptional activation domains.

Several years ago, we inadvertently observed polymerization of the LC domains of the FET proteins, as well as certain hnRNP proteins (Kato et al., 2012). When incubated at high concentration, the LC domains of these proteins polymerize into amyloid-like fibers. A combination of X-ray diffraction and electron microscopy gave evidence that the fibers were of the prototypic cross-β structure first described 50–60 years ago by Astbury et al. (1959). Unlike irreversible, pathogenic amyloids, the fibers polymerized from LC domains present in FUS, TAF15 and hnRNPA2 are readily disassembled upon dilution. By comparing the effects of mutations in the LC domains of FUS and TAF15 on both transcription activation capacity and polymerization, a strong correlative relationship gave evidence that polymerization might be of critical importance to the function of these domains in living cells (Kwon et al., 2013).

Heretofore missing from this line of research was any evidence indicative of the structure of LC domains in their native state. In attempts to address this shortcoming, we developed a chemical probing strategy that allows generation of a footprint indicative
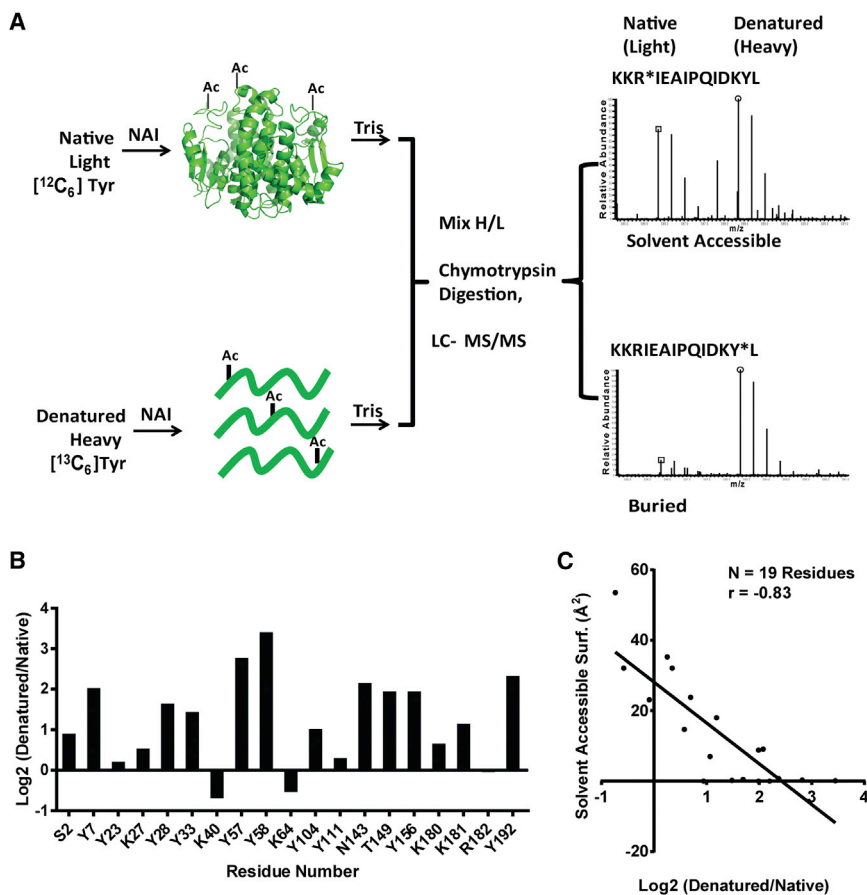
A

B

C

Figure 1. Differing Patterns of Acetylation of Folded and Denatured Samples of Glutathione-S-Transferase Mediated by N-Acetylimidazole

(A) Folded GST was exposed to NAI under conditions leading to roughly one modification per polypeptide chain, with the reaction quenched by the addition of 0.8 M Tris. A separate batch of GST grown in bacterial cells supplemented with ^{13}C-labeled tyrosine was denatured in 5 M guanidine thiocyanate prior to NAI treatment. Following quenching with Tris, the two samples were mixed, digested with chymotrypsin, and subjected to SILAC mass spectrometry.

(B) 19 acetylated side chains were scored for abundance in the two samples, yielding an NAI footprint. The degree of residue protection from NAI modification in the folded state, relative to the denatured state, is measured on the y axis as log2 values.

(C) Plot showing the correlative relationship between the degree of protection from NAI in the folded state, relative to the denatured state (x axis), and the measured level of solvent accessibility determined from the X-ray crystal structure of GST (y axis).

See also Figure S1 and Table S1 and S2.

of ordered structure. After having validated the utility of the approach using two enzymes of known structure, we deployed the footprinting strategy on fibrous polymers of the LC domain of hnRNPA2. Our observations give evidence that the LC domain of hnRNPA2 exists in the same structural state in both recombinant polymers of the protein and native hnRNPA2 within the nuclear compartment of mammalian cells.

RESULTS

Development of a Chemical Footprinting Method

N-acetylimidazole (NAI) is a reactive chemical that is capable of acetylating certain amino acid side chains in proteins (Riordan et al., 1965; Timasheff and Gorbunoff, 1967). Under conditions of neutral pH, the chemical can donate an acetyl group to serine, tyrosine, lysine, threonine, arginine, and asparagine side chains. Reasoning that the ability of NAI to modify amino acids might be influenced by the structural state of a protein, we compared modification of glutathione-S-transferase (GST) as a function of its folded versus unfolded state. GST enzyme was prepared under conditions of isotopic labeling with ^{13}C-labeled tyrosine to produce a "heavy" protein sample. This sample was denatured with a chaotropic reagent and exposed to NAI under conditions leading to roughly one modification per polypeptide. The reaction was quenched with 0.8 M Tris (pH 8.8), which inactivates

the NAI reagent. A corresponding "light" sample of GST was exposed to NAI in its folded state, again under conditions resulting in roughly one modification per polypeptide and again quenched with Tris to terminate the reaction. The heavy and light samples were mixed at a 1:1 ratio, digested with chymotrypsin and then evaluated by SILAC mass spectrometry (Figure 1A and Experimental Procedures).

The patterns of NAI reactivity with the denatured and folded states of GST were different. Certain amino acid side chains reacted similarly in the two protein samples (Y23, K27, K40, K64, Y111, and R182), whereas others were acetylated to a lesser extent in the folded sample compared with denatured GST (Y7, Y57, Y58, and Y192). The NAI "footprint" of GST is shown in Figure 1B. We then compared this footprint with the degree of surface exposure of NAI-modified side chains as deduced from the X-ray crystal structure of the enzyme (Rufer et al., 2005) (PDB: 1Y6E). A strong correlative relationship was observed between NAI accessibility and solvent exposure in the structure (Figure 1C and Table S1). Surface-exposed residues tended to be NAI accessible, whereas residues buried within the core of the enzyme tended to be NAI inaccessible. We conclude that the correlative match between NAI-accessibility and protein structure gives evidence that the NAI footprint is properly reflective of protein structure.

Proceeding from a recombinant protein sample to a native protein within mammalian nuclei, we evaluated the difference in NAI modification of the poly-ADP-ribose polymerase (PARP) enzyme as a function of its folded versus denatured state. Nuclei were prepared from 293T cells that had been grown in either normal tissue culture medium (light) or medium deprived of tyrosine and supplemented with an isotopically labeled form of the

amino acid (heavy). The light sample of nuclei was exposed to a 30 mM level of NAI for 15 min before quenching with Tris. The heavy sample was denatured in 5 M guanidine thiocyanate prior to exposure to the same level of NAI and then also quenched with Tris. The samples were combined, digested with chymotrypsin overnight, and processed by mass spectrometry (Experimental Procedures).

NAI modification was monitored on 14 amino acid side chains in the native and denatured forms of PARP (Figure S1A). Six residues were modified by NAI far more extensively in the denatured sample than the intact enzyme (K621, T799, K802, Y817, S902, and S904), five residues were modified slightly more extensively in the denatured sample relative to the intact enzyme (K571, S782, S783, S808, and K816), and three residues were modified equally in the two samples (K616, K903, and Y907). We again observed a correlation between NAI accessibility and protein structure (PDB: 3GJW) (Figure S1B and Table S2). The three side chains that were modified equally in the two samples show a high level of predicted solvent accessibility in the X-ray crystal structure of PARP (Miyashiro et al., 2009). Likewise, five of the six residues observed to be highly protected from NAI modification are predicted to be solvent inaccessible by the crystal structure of the enzyme.

Analysis of three consecutive residues in the polypeptide chain of PARP is particularly revealing. Serine residue 902 is protected from NAI modification in nuclear PARP and buried beneath the surface of the enzyme. Lysine residue 903 is surface exposed and NAI accessible in the folded form of PARP. Finally, serine residue 904 is NAI inaccessible in the folded enzyme and buried beneath the surface of the PARP crystal structure. We offer that the correlative relationship between NAI accessibility and the predicted level of surface exposure of a given amino acid side chain validates this means of probing protein structure both in a recombinant protein and a native enzyme present in nuclei of mammalian cells.

Determination of the NAI Footprint of Recombinant hnRNPA2 Fibers

Hydrogel droplets were formed using a fusion protein linking mCherry to the LC domain of hnRNPA2 (Kato et al., 2012). This protein sample was exposed to NAI under conditions resulting in roughly one modification per polypeptide chain and then quenched with Tris (Experimental Procedures). Similarly prepared hnRNPA2 polymers were formed using protein isotopically labeled with heavy tyrosine. The latter sample was denatured with guanidine thiocyanate prior to NAI-mediated modification, followed by quenching with Tris. The two samples were combined at a 1:1 ratio, digested with chymotrypsin, and then analyzed by mass spectrometry (Experimental Procedures).

23 amino acid side chains were evaluated for NAI accessibility. 12 amino acids appeared to be equally accessible to NAI-mediated modification in the 2 samples, and 11 appeared to be less accessible in the native fibers relative to the denatured protein sample (Figure 2). Three of these acetylated amino acid residues could be identified in the same peptide spanning amino acids 302–319 of the hnRNPA2 polypeptide. High-performance liquid chromatography (HPLC) chromatography was successful in separating variants of this peptide acetylated at lysine 305,

serine 306, or serine 312 (Figure 2B). The peptide variant acetylated at K305 was found at equal abundance in both light and heavy samples, indicative of the ability of NAI to modify this residue irrespective of whether the protein was in the fibrous or denatured state. The variant acetylated at S306 was considerably less abundant in the light sample than the heavy sample, giving evidence of its protection from NAI modification in the fibrous state. Finally, the variant acetylated at S312 was slightly less abundant in the light sample relative to the heavy sample, which is consistent with partial protection from NAI modification when the LC domain of hnRNPA2 existed in the polymeric state.

The pattern of protection from NAI modification in polymeric fibers of the hnRNPA2 LC domain, or lack thereof, can be described in the following way. An extensive, N-terminal region of the protein was equally acetylated by the chemical probe irrespective of the fibrous or denatured state. An equally extensive segment corresponding to a more C-terminal region of the LC domain was protected in the polymeric state, relative to the denatured state, at 11 out of 12 acetylated residues. Right within the middle of this apparently ordered region of the LC domain, lysine residue 305 was found to be equally accessible in both the polymeric and denatured states of the protein. Finally, the three most C-terminal residues scored in the assay were all equally accessible under both fibrous and denatured states.

Relationship of hnRNPA2 Footprints between Recombinant and Nuclear Forms of the Protein

Using the same methods described for determining an NAI footprint for the nuclear form of the PARP enzyme (Figure S1), we probed the structure of native hnRNPA2 present in nuclei freshly prepared from 293T cells (Experimental Procedures). Isotopically labeled heavy protein was probed under the denaturing conditions of 5 M guanidine thiocyanate. Light protein was probed via the exposure of nuclei to the NAI chemical reagent. Following quenching with Tris, the samples were mixed, digested with chymotrypsin, and evaluated by mass spectrometry.

The NAI footprint observed for native, nuclear hnRNPA2 could be scored for 18 of the 23 acetylated residues observed in the footprint derived from recombinant hnRNPA2, and the two footprints were qualitatively similar (Figure 3A). Of the acetylation events detected in both footprints, all nine residues that were equally accessible to NAI-mediated acetylation in both polymeric and denatured samples of recombinant hnRNPA2 protein were also acetylated equally in the native hnRNPA2 irrespective of structural state. Seven of the eight residues that were preferentially protected from acetylation as a function of the fibrous state of recombinant hnRNPA2 protein were also preferentially protected in the native hnRNPA2 protein relative to nuclear protein that had been denatured with 5 M guanidine thiocyanate. The single qualitative difference between the two footprints was tyrosine residue 324. This residue was preferentially protected from NAI-mediated acetylation in the fibrous form of recombinant hnRNPA2 yet was equally accessible to the chemical probe in native hnRNPA2 irrespective of whether nuclei were left intact or denatured.

Despite displaying qualitative similarities, the NAI-generated footprints for recombinant and native hnRNPA2 differed quantitatively in a consistent manner. The NAI protected residues observed in recombinant hnRNPA2 yielded an average of roughly
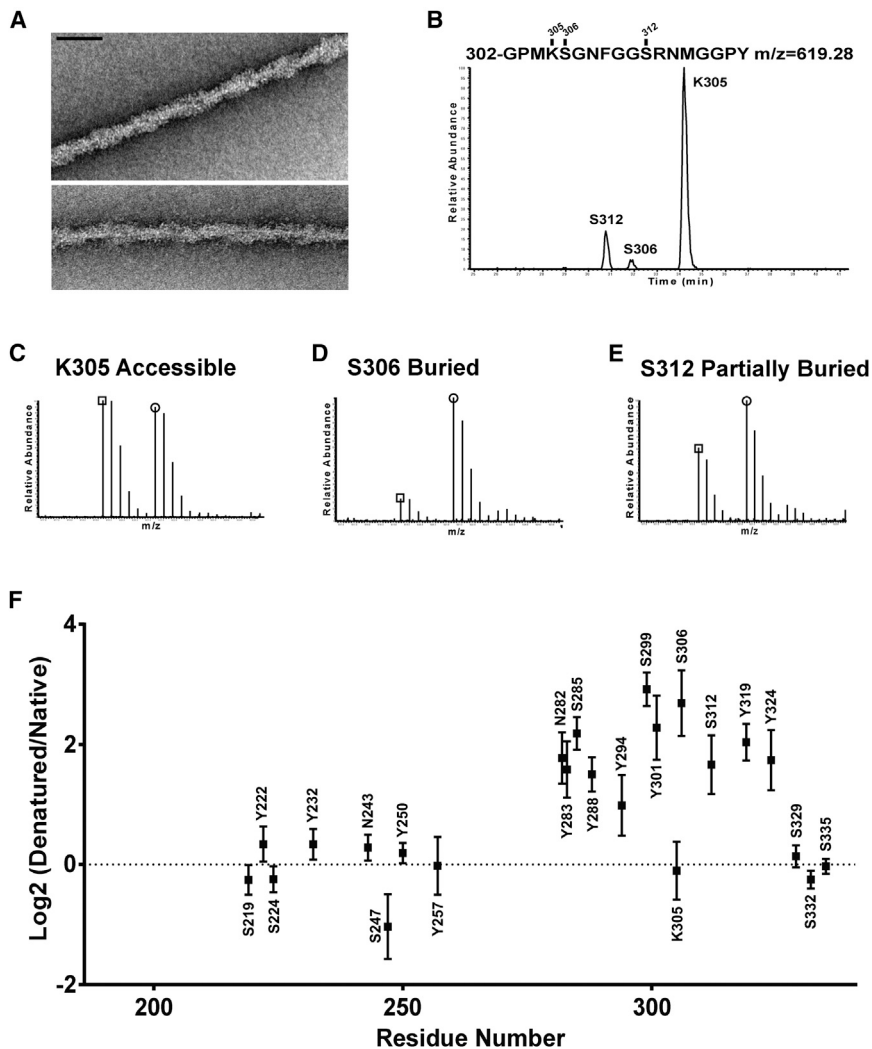
A

B

302-GPMKSGNFGGSRNMGGPY m/z=619.28

C K305 Accessible

D S306 Buried

E S312 Partially Buried

F

Figure 2. Footprint of NAI-Mediated Acetylation of Recombinant hnRNPA2 Polymeric Fibers

(A) Electron micrographs of negatively stained polymeric fibers formed from an mCherry:hnRNPA2 fusion protein (Experimental Procedures). Scale bar, 70 nm.

(B) HPLC separation of chymotryptic digestion products of the LC domain of hnRNPA2 corresponding to residues 302–319. The S312 acetylated peptide eluted earlier from the column than the S306 acetylated peptide, which, in turn, eluted earlier than the K305 acetylated peptide (Experimental Procedures).

(C) Relative abundances of the K305 acetylated peptides in folded versus denatured samples.

(D) Relative abundances of the S306 acetylated peptides in folded versus denatured samples.

(E) Relative abundances of the S312 acetylated peptides in folded versus denatured samples.

(F) NAI footprint of the LC domain of hnRNPA2 (all data are presented as means ± SD).

See also Figure S2 and Table S3.

3-fold (log2 ~1.8) difference when comparing peptide abundance in the light (fibrous) and heavy (denatured) samples. Turning to the native hnRNPA2 assayed in either intact or denatured nuclei, the average difference in peptides revealing NAI protected residues was roughly 1.5-fold (log2 ~0.5). Interpreted most simply, this difference gives indication that a smaller fraction of the native hnRNPA2 present in nuclei may exist in the structurally ordered state than the fraction deduced by studies of recombinant hnRNPA2 polymeric fibers.

Co-expression of hnRNPA2 with Peptidyl-prolyl Cis-trans Isomerase Causes Tyrosine 324 to Become NAI Accessible in Recombinant Polymers

The NAI footprint observed in recombinant hnRNPA2 polymeric fibers was qualitatively similar to that observed for native hnRNPA2 in intact nuclei. Among 18 residues defining the footprint, tyrosine 324 was the single amino acid that was clearly different in the two samples. This residue was protected from NAI-mediated acetylation in fibrous preparations of recombinant hnRNPA2, but not in the native hnRNPA2 present in intact nuclei.

Proline residues are found six positions on the amino terminal side of tyrosine 324, and two positions on its carboxyl terminal side (Figure S2). Proteomic studies of cellular proteins that bind to hydrogel droplets formed from the LC domains of both hnRNPA2 and FUS revealed retention of peptidyl-prolyl cis-trans isomerase 1 (PPIA), the most abundant isoform of a family of peptidyl-prolyl cis-trans isomerase enzymes. PPIA has been reported to interact with RNA granule proteins upon biochemical fractionation (Lauranzano et al., 2015), and antibodies to the enzyme revealed co-localization with stress granules (Figure S3). We thus reasoned that the PPIA enzyme might affect the structure of hnRNPA2 fibers by facilitating cis-trans interconversion of the peptide bonds of proline residue 319 or 326 of the hnRNPA2 polypeptide.

To test this hypothesis, mCherry:hnRNPA2 was co-expressed with either the native form of PPIA or a catalytically inactive mutant (Zydowsky et al., 1992). Following purification of the mCherry:hnRNPA2 protein, polymeric fibers were formed and exposed to the NAI probe under either the polymeric or denatured state. Co-expression of hnRNPA2 with the active form of PPIA yielded an NAI footprint wherein tyrosine residue 324 was equally accessible to acetylation irrespective of fibrous or denatured state (Figure 3B, top). By contrast, co-expression with the catalytically inactive form of PPIA yielded a footprint indistinguishable from that seen on recombinant hnRNPA2 never exposed to the enzyme (Figure 3B, bottom).

The bottom panel of Figure 3 (Figure 3C) correlatively compares the NAI footprints of hnRNPA2 observed in native protein within intact nuclei with that of recombinant protein expressed in either the absence or presence of PPIA. The r-value of correlation

Figure 3. NAI Footprints of the LC Domain of hnRNPA2 Deduced from Recombinant Protein, Native Nuclear hnRNPA2, and Recombinant Protein Co-expressed with Peptidyl-prolyl Cis-trans Isomerase

(A) NAI footprint of recombinant hnRNPA2 fibers as described in Figure 2 (upper footprint) compared with NAI footprint deduced from native, nuclear hnRNPA2 (lower footprint). Note that tyrosine 324 is protected from NAI modification in the folded, recombinant form of hnRNPA2, but not in the footprint deduced from the native, nuclear protein.

(B) NAI footprint of recombinant hnRNPA2 co-expressed with active PPIA enzyme (upper footprint) compared with footprint of hnRNPA2 co-expressed with a catalytically inactive form of the enzyme (lower footprint). Note that co-expression of hnRNPA2 with the active form of PPIA causes tyrosine 324 to become exposed to NAI modification in the polymeric state.

(C) Plots showing the correlative relationship of the NAI footprint of recombinant hnRNPA2 to that of the native, nuclear form of the protein. Correlation plot on left compares the footprint of recombinant hnRNPA2 not exposed to the PPAI enzyme with the nuclear hnRNPA2 footprint. Correlation plot on right compares the footprint of recombinant hnRNPA2 co-expressed with the active PPIA enzyme with the nuclear hnRNPA2 footprint.

See also Figure S3.

of the native and recombinant footprints was 0.76, which increased to 0.89 when the recombinant hnRNPA2 had been co-expressed with PPIA.

Mutations in the NAI-Protected Region of the hnRNPA2 LC Domain Impede Hydrogel Binding

Is the NAI footprint telling us anything of functional relevance to the LC domain of hnRNPA2? To address this question, we prepared mutated variants of the LC domain of hnRNPA2 wherein all 25 phenylalanine and tyrosine residues were individually mutated to serine (Figure S2). GFP fusion proteins representing wild-type hnRNPA2 and all of the individual mutants were expressed in bacterial cells, purified, and assayed for the ability to adhere to mCherry:hnRNPA2 hydrogel droplets (Figure 4A).

Of the 25 mutants, 6 were found to substantially impede binding to hydrogel droplets formed from mCherry fused to the wild-type LC domain of hnRNPA2. Five of the six tyrosine- or phenylalanine-to-serine mutations that substantially impede hydrogel binding occur within the region of the LC domain that is protected from NAI modification in the fibrous state (Y278S, Y283S, F291S, F309S, and Y319S). The sixth mutant that significantly impeded in hydrogel binding, Y264S, occurs on the amino terminal side of the NAI protected region within a span where we failed to find acetylated side chains—a dead zone in the footprint

(residues 258–282, Figure 2F). We tentatively conclude that these six residues are particularly important for polymerization of hnRNPA2 and that polymerization causes NAI protection.

The remaining 19 mutants fell into two categories with respect to hydrogel binding. 12 mutants bound to hydrogels in a manner indistinguishable from wild-type hnRNPA2. Two of these mutants, Y335S and Y341S, were located in the very C-terminal region of the LC domain, concordant with a small region that was fully accessible to NAI modification irrespective of whether the protein was in a polymeric or denatured state. Seven of these phenotype-void mutants, F95S, F197S, F207S, F215S, Y222S, F228S, and Y250S, were located in the amino terminal region of the LC domain that was widely accessible to NAI modification irrespective of structural state. The remaining three mutations that had no discernible effect on hydrogel binding, F244S, Y257S, Y275S, were all localized in the dead zone of the NAI footprint. Finally, seven mutants, including Y235S, Y250S, Y271S, Y288S, Y274S, Y301S, and Y324S, mildly affected binding to mCherry:hnRNPA2 hydrogels. These seven mutants mapped randomly across the LC domain of hnRNPA2. We conclude that tyrosine- and phenylalanine-to-serine mutations in NAI protected regions impede hydrogel binding, whereas those in NAI accessible regions do not impede hydrogel binding. This conclusion favors functional significance of the NAI footprint.
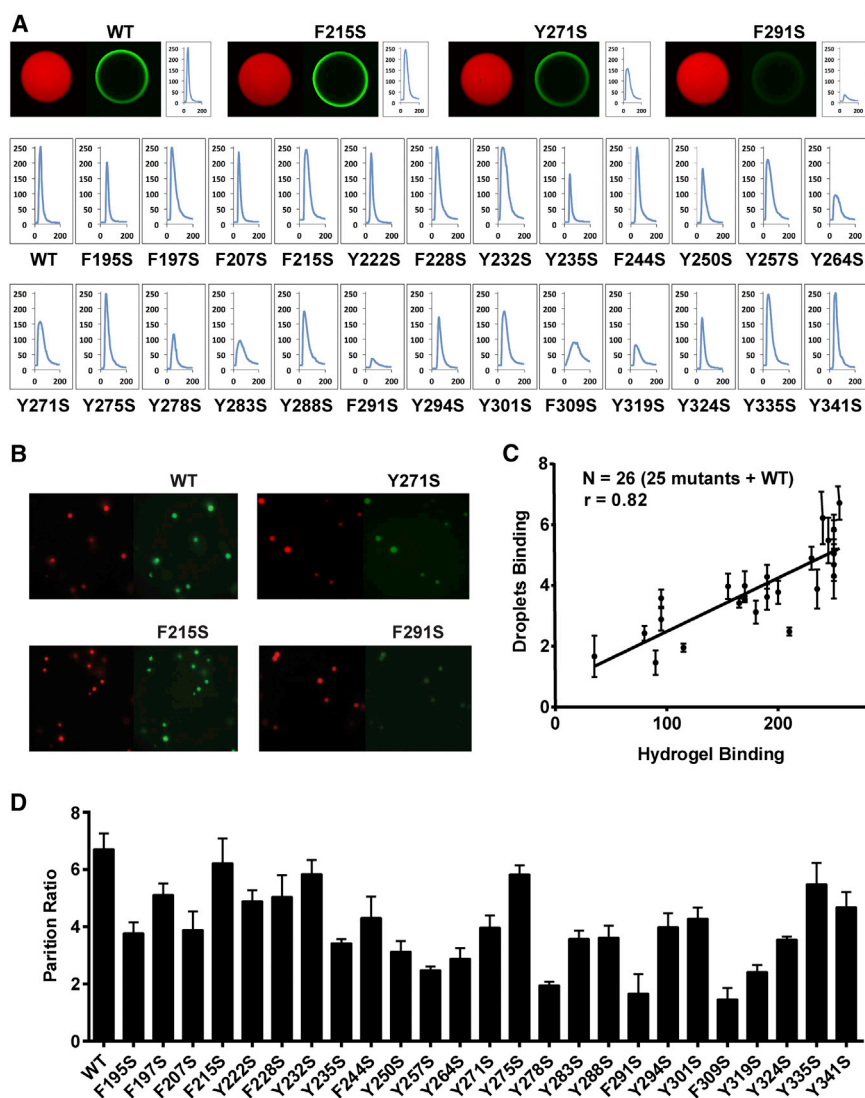
Figure 4. Correlative Relationship between Binding of Mutated Variants of the LC Domain of hnRNPA2 to Hydrogels Relative to Their Partitioning into Liquid-like Droplets

(A) All phenylalanine and tyrosine residues within the LC domain of hnRNPA2 were individually mutated to serine, expressed as GFP fusion proteins, purified and tested for binding to mCherry:hnRNPA2 hydrogel droplets (Experimental Procedures). Top figures show images of hydrogel binding by GFP linked to the native LC domain of hnRNPA2 (WT), the F215S mutant, the Y271S mutant, and the F291S mutant. Confocal images were scanned to yield the signal intensity of bound GFP (Experimental Procedures), yielding the 26 scans in the lower part of the figure. x axis indicates the scanned distance in μm, and the y axis indicates the GFP signal intensity in arbitrary units.

(B) Liquid-like droplets formed upon binding of a PTB:hnRNPA2 fusion protein to a synthetic RNA containing five copies of the PTB recognition sequence (Experimental Procedures; see also Figure S4). The presence of a SNAP tag allowed the PTB:hnRNPA2 fusion protein to be appended with a red dye. When exposed to GFP alone, no partitioning into liquid-like droplets was observed (data not shown). When exposed to GFP fused to the native LC domain of hnRNPA2 (WT), clear evidence of partitioning was observed within minutes. Certain phenylalanine- or tyrosine-to-serine mutants partitioned well into liquid-like droplets (F215S), whereas others did not (Y271 and F291S).

(C) Plot showing the correlative relationship between hydrogel binding and partitioning into liquid-like droplets for GFP linked to the native (WT) LC domain of hnRNPA2 along with 25 individual phenylalanine- and tyrosine-to-serine mutants.

(D) Partitioning into liquid-like droplets was quantified for all phenylalanine- and tyrosine-to-serine mutants that had been constructed and assayed for binding to mCherry:hnRNPA2 hydrogel droplets (A). Histogram shows relative levels of partitioning of GFP linked to the native (WT) LC domain of hnRNPA2 as compared with the 25 individual mutants.

See also Supplemental Information and Figure S2.

Mutations in the LC Domain of hnRNPA2 Act Correlatively on Hydrogel Binding and Partitioning into Liquid-like Droplets

During the preparation of GFP:FUS hydrogel droplets, we have long observed that the concentrated protein solutions become cloudy prior to gelation. Reanalysis of a His$_6$-tagged LC domain of FUS by light microscopy revealed the cloudy solution to be composed of liquid-like droplets (Figure S4). A number of investigators have recently reported that LC domains from a variety of proteins, including FUS, hnRNPA1, and DDX4, can prompt formation of liquid-like droplets (Altmeyer et al., 2015; Lin et al., 2015; Molliex et al., 2015; Nott et al., 2015; Patel et al., 2015).

It is of potential importance to know whether the physical forces leading to hydrogel formation (polymerization of LC domains) are the same or different from those leading to liquid-like droplets. To this end, we have followed the procedures of Lin et al., (2015) to create liquid-like droplets driven by the LC domain of hnRNPA2. A triple fusion protein was prepared linking the LC domain of hnRNPA2 on the C-terminal side of a polypyrimidine tract-binding protein (PTB) RNA binding domain, which was in turn linked to maltose binding protein (MBP), with a tobacco etch virus (TEV) protease cleavage site between the MBP and PTB domains (Experimental Procedures and Figure S4B). The MBP:PTB:hnRNPA2 LC domain fusion further contained a His$_6$ tag at its C terminus, as well as a SNAP tag for dye labeling on the N-terminal side of the PTB domain.

Following co-expression with PPIA, purification via nickel and amylose resin chromatography, the protein was mixed with a synthetic RNA containing five copies of a PTB binding site and exposed to TEV protease. Within 10 min, liquid-like droplets could be observed by light microscopy (Figure 4B). We then deployed a droplet partitioning assay to assess whether GFP:hnRNPA2 LC domain fusion proteins could be incorporated into the liquid-like droplets. Recombinant GFP-alone protein

A
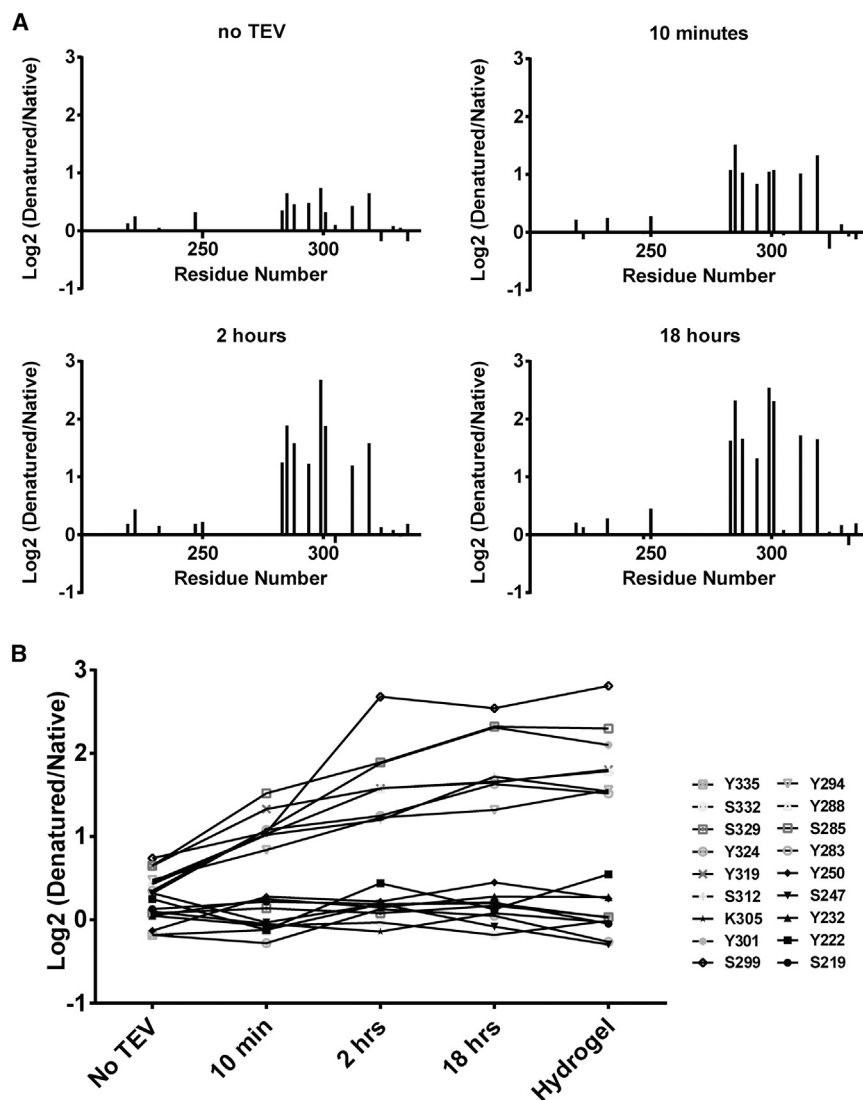
Figure 5. Liquid-like Droplets Display the Same NAI Footprint as Found in Hydrogel Polymers and the Native hnRNPA2 Present in Nuclei Freshly Isolated from Mammalian Cells

(A) A fusion protein linking maltose binding protein (MBP) to the RNA binding domains of PTB and the LC domain of hnRNPA2 (Figure S4B) was co-expressed with the peptidyl-prolyl cis-trans isomerase enzyme (PPIA), purified, and mixed with a synthetic RNA containing five PTB binding sites. Addition of TEV protease triggered the rapid formation of liquid-like droplets (Figure 4B). Protein samples were footprinted with the NAI reagent as a function of time before and after TEV protease cleavage. Hints of the NAI footprint could be seen in the protein sample before exposure to TEV protease, and the intensity of the footprint was sequentially enhanced at the 10 min, 2 hr, and 18 hr post-cleavage time points.

(B) The log2 ratio of NAI protection for all of the 18 acetylated amino acids is plotted on the y axis as a function of time post-exposure to TEV protease (x axis).

See also Table S3.

B

Y335		Y294	
S332		Y288	
S329		S285	
Y324		Y283	
Y319		Y250	
S312		S247	
K305		Y232	
Y301		Y222	
S299		S219	

Liquid-like Droplets Display the NAI Footprint Found in Hydrogel Polymers and Nuclear hnRNPA2

If the mutational effects driving hydrogel binding and liquid-like droplets correlate, it is possible that the LC domain of hnRNPA2 might adopt similar structures in both states. To address this question, we performed NAI footprinting on the LC domain of hnRNPA2 in the context of the MBP:PTB:hnRNP LC domain fusion protein before TEV cleavage, immediately upon seeing the formation of liquid-like droplets, 2 hr after droplet formation, and 18 hr after droplet formation.

As shown in Figure 5, evidence of the canonical NAI footprint on the hnRNPA2 LC domain could be detected even before TEV protease cleavage. The quantitative intensity of the footprint became sequentially enhanced at each of the later time points. Specifically, the degree of difference in NAI protected residues between native and denatured samples was—across all protected residues—most pronounced in the 18 hr sample, less so in the 2 hr sample, further reduced in the 10 min sample, and least pronounced in the sample assayed prior to TEV protease cleavage. For the 18 hr time point, the degree of protection from NAI-mediated acetylation of buried side chains was indistinguishable between liquid-like droplets and hydrogels.

Concordant with the observations of others who have studied LC domain partitioning to liquid-like droplets (Lin et al., 2015; Molliex et al., 2015; Patel et al., 2015), we conclude that, as a function of time, LC polymerization is progressively enhanced within liquid-like droplets. Petri et al. (2012) have reported similar

was not enriched in these liquid-like droplets relative to the surrounding buffer. The GFP fusion linked to the wild-type LC domain of hnRNPA2 was rapidly incorporated into liquid-like droplets. Using this assay, we evaluated all 25 mutants that had been scored for hydrogel binding (Figures 4B and 4D).

Six mutants were impeded by more than 50% with respect to partitioning into liquid like droplets (Y257S, Y264S, Y278S, F291S, F309S, and Y319S), another eight mutants were partially impeded (F195S, F207S, Y235S, Y250S, Y283S, Y288S Y294S, and Y301S), and the remaining mutants were incorporated into liquid-like droplets in a manner indistinguishable from the wild-type LC domain (Figure 4D). The correlation plot shown in Figure 4C gives evidence of a strong concordance (r = 0.83) between the effects of mutations on hydrogel binding and partitioning into liquid-like droplets. We offer that this concordance gives evidence that similar regions of the protein promote both hydrogel binding and partitioning into liquid-like droplets and that the chemical interactions that drive both processes are likely to be the same.

observations as a function of maturation of liquid-like droplets formed from FG repeats associated with nucleoporin proteins. In summary, mutational studies of the LC domain of hnRNPA2 give evidence that similar forces drive both hydrogel retention and partitioning into liquid-like droplets, and NAI footprinting studies reveal evidence that the LC domain of hnRNPA2 adopts a similar structure in both settings.

DISCUSSION

Cells display a variety of organized puncta that, unlike mitochondria, lysosomes, chloroplasts, and peroxisomes, are not membrane invested. These include various nuclear structures, including nucleoli, nuclear speckles and para-speckles, promyelocytic leukemia (PML) bodies, Cajal bodies and histone locus bodies (Mao et al., 2011). Cytoplasmic puncta include RNA granules, P-bodies, neuronal granules, stress granules, and the polar granules of fly and worm embryos that assist in determination of the germlineage (Anderson and Kedersha, 2009). Light microscopic studies of RNA granules have led to the idea that the granule components exist in a liquid-like state separated in phase from the cytoplasm (Brangwynne et al., 2009).

Studies that may be pertinent to the biochemical forces leading to the organization of these cellular structures have begun to appear over the past several years. A potentially common conceptualization may tie two orthogonal approaches together. Li et al. (2012) have provided evidence that multivalent, polymeric structures form when proteins containing repeated SRC homology 3 (SH3) domains are mixed with proteins containing repeated proline-rich motifs (PRMs). Upon heterotypic polymerization into dendritic assemblies, these proteins undergo phase separation into spherical, liquid-like droplets.

Parallel and contemporary to the Li et al. (2012) study, we have been studying the LC sequences associated with a variety of RNA binding proteins (Han et al., 2012; Kato et al., 2012). In our case, concentrated samples of these proteins have been observed to adopt a gel-like state. Reasonably clear evidence has been gathered to support the conclusion that hydrogel formation equates to polymerization of the LC sequences. Studies of hydrogels have revealed X-ray diffraction patterns consistent with cross-β structure, and electron microscopic evaluation of hydrogels has revealed homogeneous polymeric fibrils.

Of significant concern to us has been the question as to whether the polymeric structures being studied in test tube reactions are of biological relevance. Heretofore, any linkage to biological utility has been limited to correlative mutagenesis. One example of this indirect approach to biological significance has been studies of the LC domains of the FET proteins, FUS, EWS, and TAF15. All three of these paralogous proteins have amino terminal LC domains that can be translocated onto DNA binding domains as the causative event leading to human cancer. When fused to the DNA binding domain of GAL4, the LC domains of FET proteins function as transcriptional activation domains (Riggi et al., 2007). Unbiased mutagenesis of the LC domains of TAF15 and FUS have yielded scores of mutants that affect polymerization to varying degrees. When tested for their capacity to activate transcription in living cells, a strong correlative relationship was observed with polymerization capacity (Kwon et al., 2013). Mutants fully capable of polymerization activate gene expression potently, mutants mildly impeded in polymerization activate transcription to an intermediate degree, and mutants that are incapable of polymerization fail to activate transcription.

Here, we add a more direct approach to inquire whether LC domains might function in cells via the same chemistries and structures leading to LC domain polymerization in test tubes. A footprinting method was developed using NAI. This chemical acetylates amino acid side chains in a manner influenced by protein structure and can be deployed as a reagent useful both for test tube biochemistry and the probing of native protein within freshly isolated nuclei (Figures 1 and S1). Using this approach, we hereby demonstrate that the footprint of the LC domain of hnRNPA2 in recombinant polymers is highly related to the footprint observed in nuclei (Figure 3). These observations are consistent with the conclusion that the LC domain of at least some proportion of hnRNPA2 in nuclei adopts a similar cross-β structure as has been characterized with recombinant polymers.

In considering the virtues and properties of liquid-like droplets as compared with hydrogels, we offer two contrasting perspectives. It is possible that the physical forces leading to the two states are entirely different. Recent studies of the DDX4 protein and LC domains associated with nucleoporin proteins characterized by FG repeats favor the utility of chemical interactions deployed to intertwine otherwise unstructured, random coil LC domains (Nott et al., 2015; Petri et al., 2012). In the case of DDX4, π-stacking between arginine and phenylalanine residues has been highlighted as a key chemical determinant for phase separation into liquid-like droplets. These interpretations are distinct from the polymerization of LC domains into cross-β structure that we consider to be the driving force for hydrogel formation.

Here, we offer the alternative perspective that cross-β polymerization may be at the heart of formation of both hydrogels and liquid-like droplets. By constructing and studying 25 mutated variants of the LC domain of hnRNPA2, we have found mutants that affect hydrogel binding significantly, mildly, or not at all (Figure 4A). The former category of mutants mapped almost exclusively to the region of the hnRNPA2 LC domain that was NAI resistant in the polymeric state (Figure 2). When tested for partitioning into liquid-like droplets, a strong correlative relationship was observed with hydrogel binding (Figure 4C). Mutants strongly impeded in hydrogel retention partitioned poorly into liquid-like droplets, mutants partially impeded in hydrogel retention were mildly impeded from entering liquid-like droplets, and mutants that bound hydrogels as well as the wild-type LC domain of hnRNPA2 partitioned effectively into liquid-like droplets.

We likewise deployed the NAI footprinting technique to liquid-like droplets and observed the same footprint as was found in hydrogels composed of the hnRNPA2 LC domain and nuclei freshly isolated from mammalian cells (Figure 5). Although these observations do not rule out the involvement of other chemical or physical forces in the formation of liquid-like droplets, we offer the conclusion that cross-β interactions between LC domains are an important component of the forces facilitating phase separation of LC sequences into liquid-like droplets.
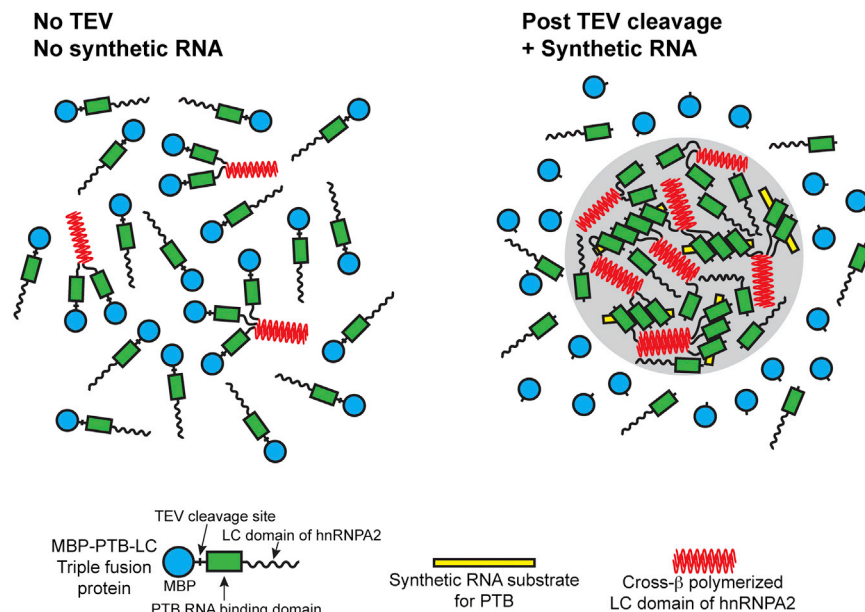
**No TEV
No synthetic RNA**

**Post TEV cleavage
+ Synthetic RNA**

MBP-PTB-LC
Triple fusion
protein

TEV cleavage site
LC domain of hnRNPA2

MBP
PTB RNA binding domain

Synthetic RNA substrate
for PTB

Cross-β polymerized
LC domain of hnRNPA2

Figure 6. Graphical Representation of Conversion of Soluble MBP:PTB:hnRNP LC Fusion Protein into Liquid-like Droplet State
The triple fusion linking maltose binding protein (MBP, blue circle), the RNA binding domain of pyrimidine track binding protein (PTB = green rectangle), and the low complexity domain of hnRNPA2 (LC domain = wavy line) remains soluble and partially polymerized via the LC domain (red sheets) prior to TEV cleavage and exposure to synthetic RNA containing five PTB binding sites (yellow rectangle). Following TEV cleavage and exposure to RNA, MBP is left in solution and PTB:hnRNP LC domain fusion protein partitions into liquid-like droplet (gray shading) in a state of enhanced polymerization.

If we adopt the simplistic interpretation that studies of hydrogels and liquid-like droplets are variations on essentially the same theme, one can consider the differences and utilities of the two systems. We reason that the sizes of polymers in hydrogels are much longer than those found in liquid-like droplets and that the size distribution and dynamics of LC domains in the latter setting may be a better representation of how LC domains function in living cells. The quantitative intensity of the NAI footprint in the various settings deployed in this study may be instructive in this regard. In hydrogel samples, the degree of NAI protection in ordered regions of the protein was roughly 3× that of denatured samples. In cells, the quantitative degree of protection was roughly 1.5×. The NAI footprint observed in freshly prepared liquid-like droplets yielded a quantitative degree of protection more closely matching to that of native hnRNPA2 as probed in isolated nuclei.

Paradoxically, evidence of the existence of a low level of cross-β structure was seen in samples of the MBP:PTB:hnRNPA2 LC domain triple fusion protein before TEV release of MBP, before exposure to the synthetic RNA containing iterative PTB binding sites, and before formation of liquid-like droplets (Figure 5). As recently articulated by Arosio et al. (2015), the initial nucleation of amyloid fibers is triggered within micro-seconds of protein mixing during the lag phase of polymerization. Transition from lag to growth phase of amyloid polymerization reflects, of course, a profound enhancement of the proportion of molecules existing in the fibrous state. We interpret the observation of an NAI footprint in samples of the triple fusion protein before TEV cleavage and exposure to the synthetic RNA substrate, and before the formation of liquid-like droplets, to reflect the same phenomenon of fiber nucleation observed during the lag phase of polymerization of pathogenic amyloid fibers. This interpretation is displayed graphically in Figure 6.

We have never thought or contended that LC polymers thousands of subunits in length are operative in living cells.

Indeed, LC domains are a cellular sink for post-translational modification, including phosphorylation, acetylation, methylation, glycosylation, and PARylation (Choudhary et al., 2009; Lee, 2012; Zhang et al., 2013). Knowing that phosphorylation can regulate the polymerization of LC domains (Han et al., 2012), we have every reason to believe that the behavior of LC domain polymers will be far more dynamic in living cells than in the hydrogels we have been studying for the past several years.

Despite recognizing hydrogels as being aberrantly static, they have offered a number of useful advantages. They have allowed us to probe for structure—first and foremost telling us that LC domain polymerization is at the heart of hydrogel formation (Kato et al., 2012). Second, they have given us assays to probe, in an unbiased manner, for both proteins and RNAs that bind to hydrogels (Han et al., 2012). Third, they have allowed us to conduct correlative mutagenesis experiments in search of mutations that affect hydrogel binding as compared with other cellular activities (Kwon et al., 2013). Fourth, they have allowed us to study interaction with LC domains that—on their own —cannot polymerize. These include the C-terminal domain of RNA polymerase II (CTD) of RNA polymerase II and the serine arginine repeat (SR) domains of pre-mRNA splicing factors, both of which bind specifically to certain hydrogels in a manner regulated by the protein kinase enzymes known to control CTD and SR domain function (Kwon et al., 2013, 2014). Finally, in this report, we have employed hydrogels to develop the NAI footprinting strategy.

Since the submission of this manuscript, four new papers have been published concerning the partitioning of LC domains into liquid-like droplets. Two of the four papers conclude that there is no biologic or physiologic role for cross-β polymerization of these LC domains and that polymerization is solely reflective of a pathologic state (Altmeyer et al., 2015; Patel et al., 2015). The two other papers conclude that cross-β polymerization of LC domains is not the driving force leading to the formation of liquid-like droplets but that it may be of biologic utility during the maturation of liquid-like droplets and/or RNA granules (Lin et al., 2015; Molliex et al., 2015). These four papers concur with the work of Elbaum-Garfinkle et al. (2015) and Nott et al.

(2015) indicating that the primary biologic utility of LC domains is driven by forces other than cross-β polymerization, perhaps including π-stacking of arginine and phenylalanine residues or other forms of weak or "fuzzy" interactions involving unfolded polypeptide domains.

Here, we offer the very different perspective that cross-β polymerization commonly drives the formation of hydrogels, the retention of LC domains trapped by hydrogels, the formation of liquid-like droplets, the partitioning of LC domains into existing liquid-like droplets, and the formation and maturation of RNA granules. In other words, we submit the hypothesis that the involvement of LC domains in the formation of RNA granules, liquid-like droplets, and hydrogels all rely on one in the same phenomenon—cross-β polymerization. Further experimentation, including derivation of the molecular structure of LC domains existing in the labile, polymeric state, should help resolve this controversy.

EXPERIMENTAL PROCEDURES

Detailed experimental procedures are available in the Supplemental Information.

Materials
N-acetylimidazole was purchased from Sigma-Aldrich (USA). Ring $^{13}C_6$-tyrosine was purchased from Cambridge Isotope Laboratories (USA). The parental vector for expression of the triple fusion protein of MBP:PTB:hnRNPA2 LC domain was provided by Dr. Michael Rosen of University of Texas Southwestern Medical Center.

Preparation of Fusion Proteins
His_6:GFP or His_6:mCherry linked to the LC domain of wild-type hnRNPA2 (residues 181–341) was overexpressed alone or co-expressed with human PPIA and purified as described previously (Kato et al., 2012). Tyrosine-to-serine mutants of GFP:hnRNPA2 LC were purified in the presence of 2 M guanidine hydrochloride.

Preparation of Heavy Proteins
The stable-isotope-labeled (heavy) proteins (His_6:hnRNPA2 LC domain and His_6:GST) were prepared with ring $^{13}C_6$-tyrosine (labeled with ^{13}C on the six carbons of the phenyl ring) by following published procedures (Baxa et al., 2007).

Acetylation of Recombinant Proteins
To acetylate denatured (heavy) proteins, the proteins were denatured by 5 M GuSCN, and acetylation reactions were carried out with 30 mM NAI and 1 mg/ml proteins. The reactions were quenched by 0.8 M Tris-HCl (pH 8.8). The light proteins were acetylated in native conditions (without GuSCN). The native and denatured proteins were mixed at a 1:1 ratio, digested by chymotrypsin and analyzed by mass spectrometry.

Acetylation of Nuclei
293T cells were cultured in DMEM high glucose media containing light or $^{13}C_6$-tyrosine, respectively. Intact nuclei from heavy or light cells were purified in hypotonic buffer and washed with beta mercaptoethanol (BME) free buffer. Light intact nuclei were resuspended in a nuclei buffer, and heavy nuclei were denatured in the nuclei buffer with 5 M GuSCN. Both samples were acetylated with 30 mM NAI at RT for 15 min, quenched by Tris, and mixed together at a 1:1 ratio. The mixture was digested by chymotrypsin and then analyzed by mass spectrometry.

Acetylation of Liquid-like Droplets
Liquid-like droplets of the MBP:PTB:hnRNPA2 LC triple fusion protein were prepared as described (Lin et al., 2015). For the native sample, NAI (30 mM)

was added to the protein solution before TEV cleavage or after TEV cleavage at the indicated time points. After incubation for 15 min at room temperature (RT), the reaction was quenched by Tris. Acetylated His_6-tagged hnRNPA2 LC (heavy) protein was used for the denatured sample. The two samples were mixed, digested by chymotrypsin, and then analyzed by mass spectrometry.

Recruitment Assays with Hydrogels and Liquid-like Droplets
Hydrogel droplets of mCherry:LC domain of wild-type hnRNPA2 were prepared as described (Kato et al., 2012). GFP:hnRNPA2 LC wild-type or mutant proteins were diluted to 1 μM in 1 ml of a gelation buffer and pipetted into the hydrogel dish. After overnight incubation at 4°C, horizontal sections of the hydrogel droplets were scanned at both the mCherry and GFP excitation wavelengths by a confocal microscope. GFP signals at a boundary area of the hydrogel droplets were scanned by the program ImageJ (Schneider et al., 2012). Liquid-like droplets formed from MBP:PTB:hnRNPA2 LC and the synthetic RNA substrate were incubated with 0.1 μM of GFP:hnRNPA2 LC wild-type or mutant domains. The droplets were deposited on a cover slide and imaged by a fluorescent microscope. GFP signals inside and outside of the liquid-like droplets were measured by the program ImageJ. The partition ratio of GFP:hnRNPA2 proteins was calculated by dividing the signal inside the droplet by the signal outside.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.040.

REFERENCES

Altmeyer, M., Neelsen, K.J., Teloni, F., Pozdnyakova, I., Pellegrino, S., Grøfte, M., Rask, M.-B.D., Streicher, W., Jungmichel, S., Nielsen, M.L., and Lukas, J. (2015). Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). Nat. Commun. 6, 8088.

Anderson, P., and Kedersha, N. (2009). RNA granules: post-transcriptional and epigenetic modulators of gene expression. Nat. Rev. Mol. Cell Biol. 10, 430–436.

Arosio, P., Knowles, T.P., and Linse, S. (2015). On the lag phase in amyloid fibril formation. Phys. Chem. Chem. Phys. 17, 7606–7618.

Astbury, W.T., Beighton, E., and Parker, K.D. (1959). The cross-beta configuration in supercontracted proteins. Biochim. Biophys. Acta 35, 17–25.

Baxa, U., Wickner, R.B., Steven, A.C., Anderson, D.E., Marekov, L.N., Yau, W.M., and Tycko, R. (2007). Characterization of beta-sheet structure in Ure2p1-89 yeast prion fibrils by solid-state nuclear magnetic resonance. Biochemistry 46, 13149–13162.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A.A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science *324*, 1729–1732.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science *325*, 834–840.

Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C.C.-H., Eckmann, C.R., Myong, S., and Brangwynne, C.P. (2015). The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. Proc. Natl. Acad. Sci. USA *112*, 7189–7194.

Han, T.W., Kato, M., Xie, S., Wu, L.C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G., and McKnight, S.L. (2012). Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell *149*, 768–779.

Hope, I.A., Mahadevan, S., and Struhl, K. (1988). Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. Nature *333*, 635–640.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., et al. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753–767.

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. Cell *155*, 1049–1060.

Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S.L. (2014). Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science *345*, 1139–1145.

Lauranzano, E., Pozzi, S., Pasetto, L., Stucchi, R., Massignan, T., Paolella, K., Mombrini, M., Nardo, G., Lunetta, C., Corbo, M., et al. (2015). Peptidylprolyl isomerase A governs TARDBP function and assembly in heterogeneous nuclear ribonucleoprotein complexes. Brain *138*, 974–991.

Lee, E.K. (2012). Post-translational modifications of RNA-binding proteins and their roles in RNA granules. Curr. Protein Pept. Sci. *13*, 331–336.

Li, P., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., et al. (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336–340.

Lin, Y., Protter, D.S., Rosen, M.K., and Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Mol. Cell *60*, 208–219.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479–490.

Manley, J.L., and Tacke, R. (1996). SR proteins and splicing control. Genes Dev. *10*, 1569–1579.

Mao, Y.S., Zhang, B., and Spector, D.L. (2011). Biogenesis and function of nuclear bodies. Trends Genet. *27*, 295–306.

Miyashiro, J., Woods, K.W., Park, C.H., Liu, X., Shi, Y., Johnson, E.F., Bouska, J.J., Olson, A.M., Luo, Y., Fry, E.H., et al. (2009). Synthesis and SAR of novel tricyclic quinoxalinone inhibitors of poly(ADP-ribose)polymerase-1 (PARP-1). Bioorg. Med. Chem. Lett. *19*, 4050–4054.

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. Cell *163*, 123–133.

Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D., and Baldwin, A.J. (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Mol. Cell *57*, 936–947.

Pabo, C.O., and Sauer, R.T. (1992). Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. *61*, 1053–1095.

Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M., et al. (2015). A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. Cell *162*, 1066–1077.

Petri, M., Frey, S., Menzel, A., Görlich, D., and Techert, S. (2012). Structural characterization of nanoscale meshworks within a nucleoporin FG hydrogel. Biomacromolecules *13*, 1882–1889.

Riggi, N., Cironi, L., Suvà, M.L., and Stamenkovic, I. (2007). Sarcomas: genetics, signalling, and cellular origins. Part 1: The fellowship of TET. J. Pathol. *213*, 4–20.

Riordan, J.F., Wacker, W.E.C., and Vallee, B.L. (1965). N-acetylimidazole: A reagent for determination of "free" tyrosyl residues of proteins. Biochemistry *4*, 1758–1765.

Rufer, A.C., Thiebach, L., Baer, K., Klein, H.W., and Hennig, M. (2005). X-ray structure of glutathione S-transferase from Schistosoma japonicum in a new crystal form reveals flexibility of the substrate-binding site. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. *61*, 263–265.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Sigler, P.B. (1988). Transcriptional activation. Acid blobs and negative noodles. Nature *333*, 210–212.

Timasheff, S.N., and Gorbunoff, M.J. (1967). Conformation of proteins. Annu. Rev. Biochem. *36*, 13–54.

Triezenberg, S.J. (1995). Structure and function of transcriptional activation domains. Curr. Opin. Genet. Dev. *5*, 190–196.

Zhang, Y., Wang, J., Ding, M., and Yu, Y. (2013). Site-specific characterization of the Asp- and Glu-ADP-ribosylated proteome. Nat. Methods *10*, 981–984.

Zydowsky, L.D., Etzkorn, F.A., Chang, H.Y., Ferguson, S.B., Stolz, L.A., Ho, S.I., and Walsh, C.T. (1992). Active site mutants of human cyclophilin A separate peptidyl-prolyl isomerase activity from cyclosporin A binding and calcineurin inhibition. Protein Sci. *1*, 1092–1099.

Cell

Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems

Graphical Abstract
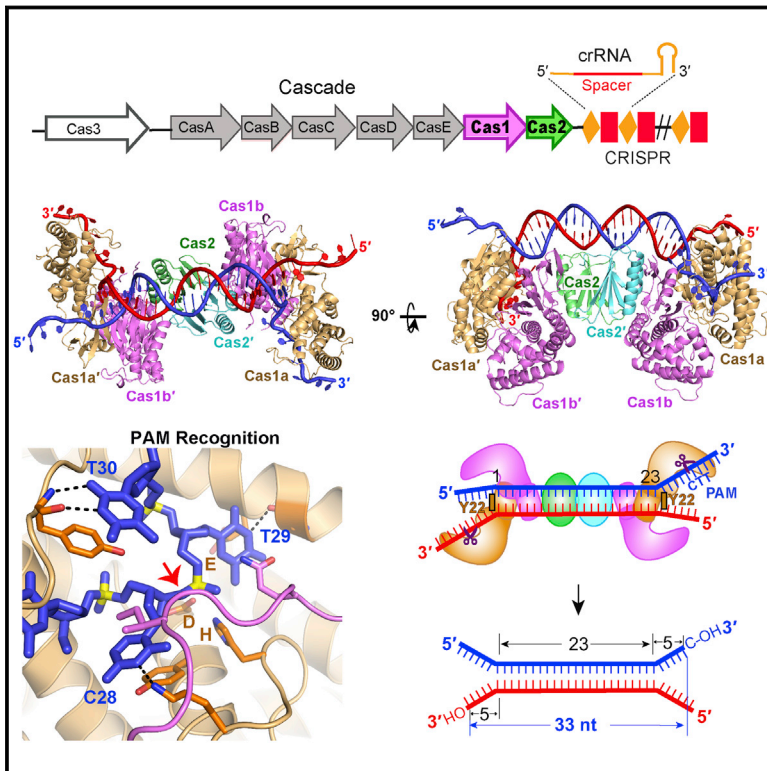
Authors

Jiuyu Wang, Jiazhi Li, Hongtu Zhao, Gang Sheng, Min Wang, Maolu Yin, Yanli Wang

Correspondence

ylwang@ibp.ac.cn

In Brief

Cas1 and Cas2 select an invading DNA sequence, termed protospacer, for insertion into the CRISPR locus of the host cell. The structure of the Cas1-Cas2-protospacer DNA complex reveals the dual-forked nature of the protospacer, explains how the protospacer is selected, and identifies how protospacer length is predetermined.

Highlights

- The dual-forked protospacer is integrated via a cut-and-paste mechanism

- Architecture of Cas1-Cas2 predetermines length of newly acquired spacer

- Cas1a recognizes PAM-complementary sequence via sequence-specific interactions

- Cas1-Cas2 undergoes a conformational change upon protospacer DNA binding

Accession Numbers

5DQU
5DLJ
5DQT
5DQZ

CellPress

Article

Cell

Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems

Jiuyu Wang,[1,2,4] Jiazhi Li,[1,2,3,4] Hongtu Zhao,[1,2,3] Gang Sheng,[1,2] Min Wang,[1,2] Maolu Yin,[1,2,3] and Yanli Wang[1,2,*]
[1]Key Laboratory of RNA Biology
[2]Beijing Key Laboratory of Noncoding RNA
Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China
[4]Co-first author
*Correspondence: ylwang@ibp.ac.cn
http://dx.doi.org/10.1016/j.cell.2015.10.008

SUMMARY

Bacteria acquire memory of viral invaders by incorporating invasive DNA sequence elements into the host CRISPR locus, generating a new spacer within the CRISPR array. We report on the structures of Cas1-Cas2-dual-forked DNA complexes in an effort toward understanding how the protospacer is sampled prior to insertion into the CRISPR locus. Our study reveals a protospacer DNA comprising a 23-bp duplex bracketed by tyrosine residues, together with anchored flanking 3′ overhang segments. The PAM-complementary sequence in the 3′ overhang is recognized by the Cas1a catalytic subunits in a base-specific manner, and subsequent cleavage at positions 5 nt from the duplex boundary generates a 33-nt DNA intermediate that is incorporated into the CRISPR array via a cut-and-paste mechanism. Upon protospacer binding, Cas1-Cas2 undergoes a significant conformational change, generating a flat surface conducive to proper protospacer recognition. Here, our study provides important structure-based mechanistic insights into PAM-dependent spacer acquisition.

INTRODUCTION

The clustered regularly interspaced short palindromic repeats (CRISPR), together with CRISPR-associated (Cas) proteins, form the microbial adaptive immune system that protects against invading phages and plasmids. The CRISPR array consists of identical short repeats interspaced by similarly sized variable spacers, which are acquired from the foreign DNA (Figure 1A) (Barrangou et al., 2007; Barrangou and Marraffini, 2014; Brouns et al., 2008). An A-T-rich leader sequence located upstream of the first repeat is essential for spacer acquisition (Yosef et al., 2012) and promotes the transcription of the CRIPSPR array (Pougach et al., 2010). The CRISPR-Cas system defends against invasive nucleic acids from phages or plasmids in three steps (van der Oost et al., 2014). First, in the spacer acquisition step (also called adaptation), a new spacer is acquired from the

invader DNA and integrated into the CRISPR locus (Barrangou et al., 2007; Fineran and Charpentier, 2012). Second, the CRISPR locus is transcribed and processed into short mature CRISPR RNA (crRNA), which then binds to Cas proteins and forms a protein-RNA complex (Brouns et al., 2008). Finally, the invading nucleic acid complementary to crRNA is recognized and degraded by the protein-crRNA complex (Garneau et al., 2010; Hale et al., 2009; Marraffini and Sontheimer, 2008). While the molecular mechanisms of expression and interference steps are now well characterized in molecular and functional terms, the adaptation step still awaits detailed analysis.

Recent studies have shown that the protospacer-adjacent motif (PAM) is fundamental to avoid auto-immunity. Only if the invading DNA is flanked by the correct PAM can it be cleaved during interference (Deveau et al., 2008). Furthermore, it was shown that PAMs are of critical importance for recognition and selection of protospacer during acquisition. It was found that protospacers flanked by the correct PAM could be incorporated into the CRISPR array (Horvath et al., 2008; Mojica et al., 2009). Interestingly, in *Escherichia coli*, the last nucleotide of the new repeat is derived from the first nucleotide of the incoming spacer, and this nucleotide is indeed the last nucleotide of the PAM sequence (Datsenko et al., 2012).

Cas1 and Cas2 are the only two Cas proteins universally conserved across all CRISPR-Cas systems (Makarova et al., 2011). Previous in vitro analysis showed that Cas1 is a metal-dependent DNase, capable of cleaving single-stranded (ss) DNA, double-stranded (ds) DNA, cruciform DNA, and branched DNA in a sequence-independent manner (Babu et al., 2011; Wiedenheft et al., 2009). Likewise, Cas2 was identified as a metal-dependent endoribonuclease that cleaves ssRNA or dsDNA (Beloglazova et al., 2008; Gunderson et al., 2015; Ka et al., 2014; Nam et al., 2012) or, alternately, shows no significant nuclease activity (Samai et al., 2010). However, one recent study demonstrated that the "active site" of Cas2 is not required for spacer acquisition (Nuñez et al., 2014), suggesting that Cas2 could play other as-yet unknown functions.

Overexpression of *E. coli* Cas1 and Cas2 induces new spacer acquisition by inserting exactly 33 nt foreign DNA behind the first repeat, indicating that Cas1 and Cas2 are both necessary and sufficient for new spacer acquisition. Previous studies demonstrated that Cas1 and Cas2 form a stable complex, which functions as an integrase that incorporates the

Cell
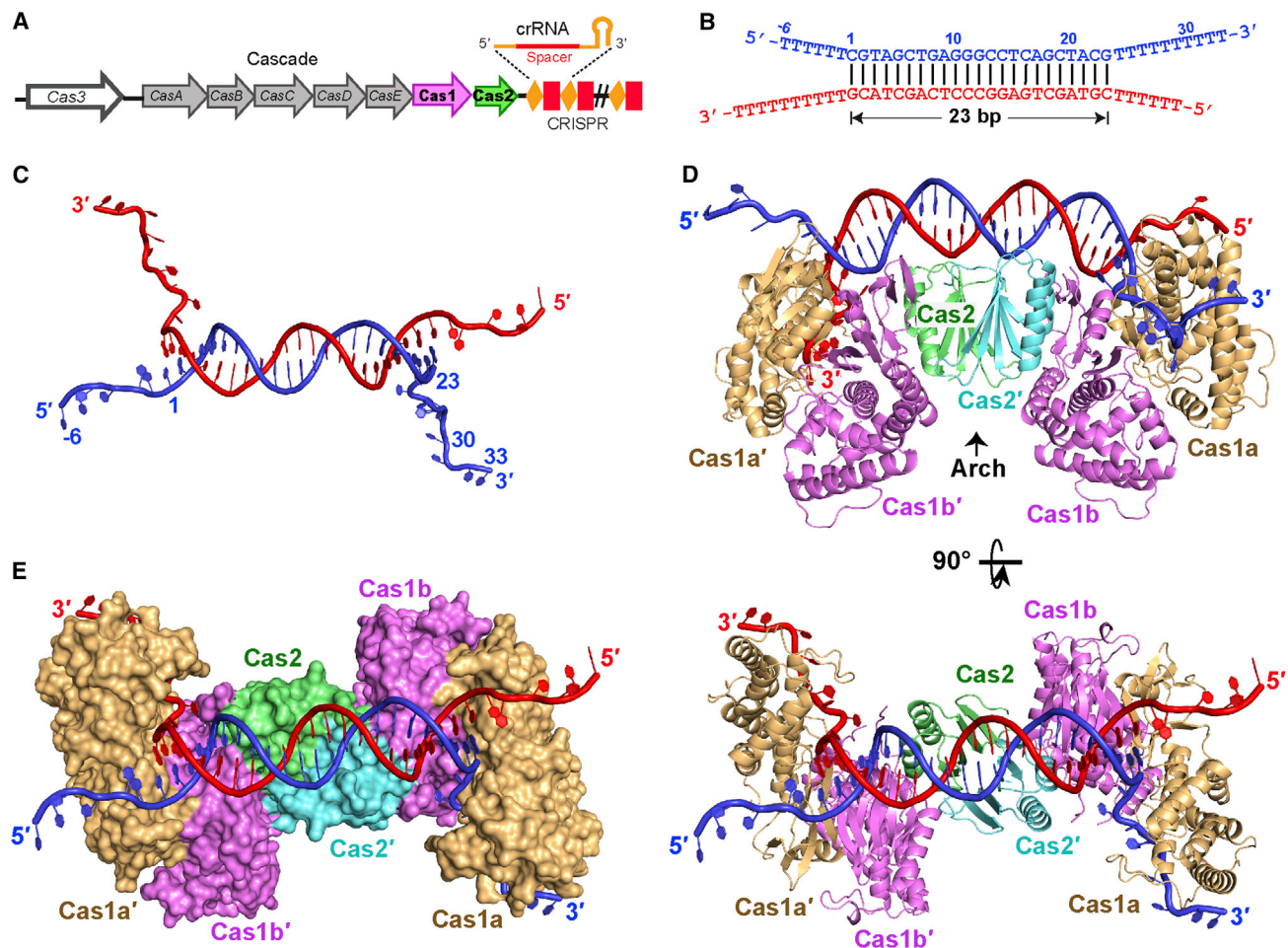
Figure 1. Crystal Structure (2.6 Å) of *E. coli* Cas1-Cas2 Bound to a Dual-Forked DNA

(A) A representation of the CRISPR-Cas locus of *E. coli* K12. The CRISPR locus consists of series of repeats (orange diamonds) that are separated by spacer sequences (red rectangles) of constant length. Cas1 and Cas2 are shown in magenta and green colors, respectively.

(B) Schematic diagram of the dual-forked DNA, which is a 23-mer palindromic duplex with $5'$-$(T)_6$ and $3'$-$(T)_{10}$ overhangs on both ends. The nucleotides in the $5'$ overhangs are numbered from −6 to −1; those in the DNA duplex are numbered from 1 to 23; and those in the $3'$ overhang are numbered from 24 to 33. The two strands of DNA are colored in red and blue, respectively.

(C) Structure of the dual-forked DNA in the Cas1-Cas2 complex.

(D) Orthogonal views of the crystal structure of the complex of Cas1-Cas2 bound to the dual-forked DNA. The Cas1a and Cas1a′ are shown in light orange, and Cas1b and Cas1b′ are show in magenta. Two monomers of Cas2 are in green and cyan, respectively. The proposed Arch segment is labeled.

(E) The surface view of the Cas1-Cas2 dual-forked DNA complex in the same orientation as Figure 1D, bottom.

See also Figure S1 and Table S1.

new spacers into the CRISPR locus (Arslan et al., 2014; Nuñez et al., 2014, 2015; Rollie et al., 2015). In *E. coli*, the integration process involves the staggered cleavage of the first CRISPR repeat, and new spacers are incorporated proximal to the leader sequence (Yosef et al., 2012). From this, three fundamental questions arise as to how Cas1-Cas2 mediates the spacer acquisition. First, what are the physiological DNA substrates of Cas1-Cas2, and what are the respective roles of Cas1 and Cas2 proteins? Second, while the spacers are known to be of a set length in each species, what are the molecular mechanisms underlying spacer length determination? Third, how does the acquisition machinery select protospacers containing a PAM sequence?

To understand the molecular mechanisms of spacer acquisition, we determined the crystal structure of *E. coli* Cas1-Cas2 bound with dual-forked DNA. Our structure highlights the following mechanistic principles related to new spacer acquisition. We demonstrate that the protospacer DNA captured by Cas1-Cas2 adopts a dual-forked form, with the $3'$ overhangs of the protospacer essential for new spacer acquisition. The PAM-complementary sequence ($5'$-CTT-$3'$), located within the $3'$ overhang, is recognized in a sequence-specific manner and is cleaved by Cas1a, generating a DNA intermediate that has 5-nt $3'$ overhangs on the two partner strands. Given that tyrosine residues cap either end of a 23-bp duplex, Cas1-Cas2 predetermines the length of the newly acquired spacer,

Cell

thereby highlighting the role of both Cas1 and Cas2 in the acquisition mechanism. Moreover, Cas1-Cas2 undergoes a significant conformational change upon protospacer binding, thereby generating optimal protospacer and target binding sites.

RESULTS

Crystal Structure of Cas1-Cas2 Bound to Single-Forked DNA

Both Cas1 and Cas2 are capable of cleaving various types of DNA in vitro. However, the exact DNA substrate of the Cas1-Cas2 in vivo has remained unknown. To obtain a crystal of the Cas1-Cas2-DNA complex, we co-crystallized the protein complex with various DNAs. As shown in Figure S1A, initially only the single-forked DNA containing a 10-bp duplex and 3′ and 5′ oligo-T overhangs of 10-nt length crystallized, resulting in a low-resolution structure of this complex at 4.5 Å.

Search and Optimization of the DNA Substrate

In terms of nomenclature, within each symmetric half of the complex, the proteins are labeled Cas1a, Cas1b, and Cas2 and Cas1a′, Cas1b′, and Cas2′. Analysis of our structures showed that this complex contains a pair of Cas1 dimers sandwiching one Cas2 dimer (Figure S1B), similar to the structure of DNA-free Cas1-Cas2 (Nuñez et al., 2014). In this 2-fold symmetric complex, the two single-forked DNAs lie on the surface of the Cas1-Cas2 in a head-to-head orientation. Each 10-bp duplex lies on the interface of a Cas1a/b dimer, with the fork facing toward the edge of the Cas1a/b dimer and the duplex end positioned on the Cas1-Cas2 interface. These findings strongly indicate that the two DNA forks always face toward the outside of Cas1-Cas2, suggesting that this orientation of the forks is fixed in the protein complex.

While the two forks are facing outward, the blunt ends of both duplexes extend toward the center, where the Cas2 dimer is located. Interestingly, the blunt ends do not meet but leave a gap in between, indicating that Cas1-Cas2 associates with duplex DNA longer than 20 bp. To test this assumption, we used various substrates, including single-fork DNA containing either 11- or 12-bp duplexes and dual-forked DNA with duplexes of 21–24 bp in length, flanked by 3′ and 5′ overhangs at both ends. To our surprise, the complex with dual-forked DNA substrates resulted in crystals with greatly improved diffraction, from which we obtained a structure of the complex at a higher resolution of 2.6 Å. This result suggests that this dual-forked DNA is closely related to the in vivo substrate used by Cas1-Cas2.

Dual-Forked DNA Is the Substrate of Cas1-Cas2

Having found a DNA substrate yielding a high-resolution structure of the complex, we found that a dual-forked DNA substrate of 23-bp duplex length flanked by 3′-terminal $(T)_{10}$ and 5′-terminal $(T)_6$ overhangs (Figure 1B) gave crystals that diffracted to the highest resolution. The structure of the complex was refined at an R_{work}/R_{free} of 0.179 and 0.207 (Table S1). The asymmetric unit contains one Cas1-Cas2-DNA complex, which possesses a pair of asymmetric Cas1 dimers (Cas1a/b and Cas1a′/b′) and

one symmetric Cas2 dimer, together with one dual-forked DNA substrate (Figures 1C, 1D, and S1C). The entire Cas1-Cas2-DNA complex exhibits 2-fold symmetry, with each half composed of Cas1a, Cas1b, and Cas2 subunits and bound DNA substrate.

In detail, the pair of symmetric Cas2 subunits are sandwiched between the pair of asymmetric Cas1 dimers (Figure 1D), similar to the single-forked DNA-bound Cas1-Cas2 complex (Figure S1B). The Cas1a/b dimer is structurally similar to its symmetry-related Cas1a′/b′ dimer counterpart, with Cas1a being similar to Cas1a′ and Cas1b similar to Cas1b′. Cas1-Cas2 is shaped like a wings-down butterfly, containing one flat top surface and an arch-shaped surface on the opposite face (Figures 1D, top, and S1D). In our structure, 14 amino acids at the N-terminal tails of Cas1a and Cas1a′ and ~40 amino acids at the C-terminal tails in both Cas1 subunits were disordered.

Within our crystal structure of the complex, the designed DNA features visible forks at either end, with a 23-bp duplex sandwiched between fork elements. The dual-forked DNA lies on the flat surface of Cas1-Cas2, and the two 3′ overhangs thread into the C-terminal domains of Cas1a and Cas1a′, respectively (Figure 1E). We observe a multitude of intermolecular interactions between the 3′ overhangs and the protein, further indicating that the dual-forked DNA is a robust substrate for the cleavage reaction by Cas1-Cas2, as discussed further below.

The DNA Duplex Segment Slots into the Flat Surface Provided by Cas1-Cas2

Next, we investigated the interaction between the DNA and the protein in the complex in greater detail. The 23-bp duplex closely follows the contours of the flat surface at the top of Cas1-Cas2, starting from Cas1a′/b′ at one end, reaching across to Cas1a/b at the other end, and interacting with intervening Cas2 along its path (Figure 1E). Comparison of the duplex in the dual-forked DNA with the canonical B-form duplex DNA shows that the interaction between the duplex and Cas1-Cas2 induces bending of the DNA (Figure S2A). As shown in Figure 2A, either end of the duplex straddles the Cas1 dimer interface. In this region, the duplex forms hydrogen bonds via its phosphate groups with Arg59, Arg245, and Arg248 of Cas1a′ and Val27, Asp29, Gly30, and Ser61 of Cas1b′ (Figures 2B and 2C). The last four base pairs (positions 19–23) of the duplex segment are stabilized by the Cas1a/b dimer in a similar manner to that observed for the symmetry-related first four base pairs (positions 1–4).

The central segment of the duplex lies on the surface of the Cas2 dimer and is stabilized by charge-charge interactions via its phosphate backbone with the positively charged Cas2 surface (Figure S2B). As shown in Figure 2D, the side chains involved in these interactions are from Arg14, Arg16, Arg77, and Arg78, together with the main chain of Asn10. Individual substitutions of these Arg residues by Ala and the double mutant of Arg77Ala and Arg78Ala reduced spacer acquisition. In addition, no new spacer acquisition was observed for Arg14Ala and Arg16Ala dual mutant (Figure 2E). Together, these results indicate that the interactions between Cas2 and duplex DNA are crucial for spacer acquisition.
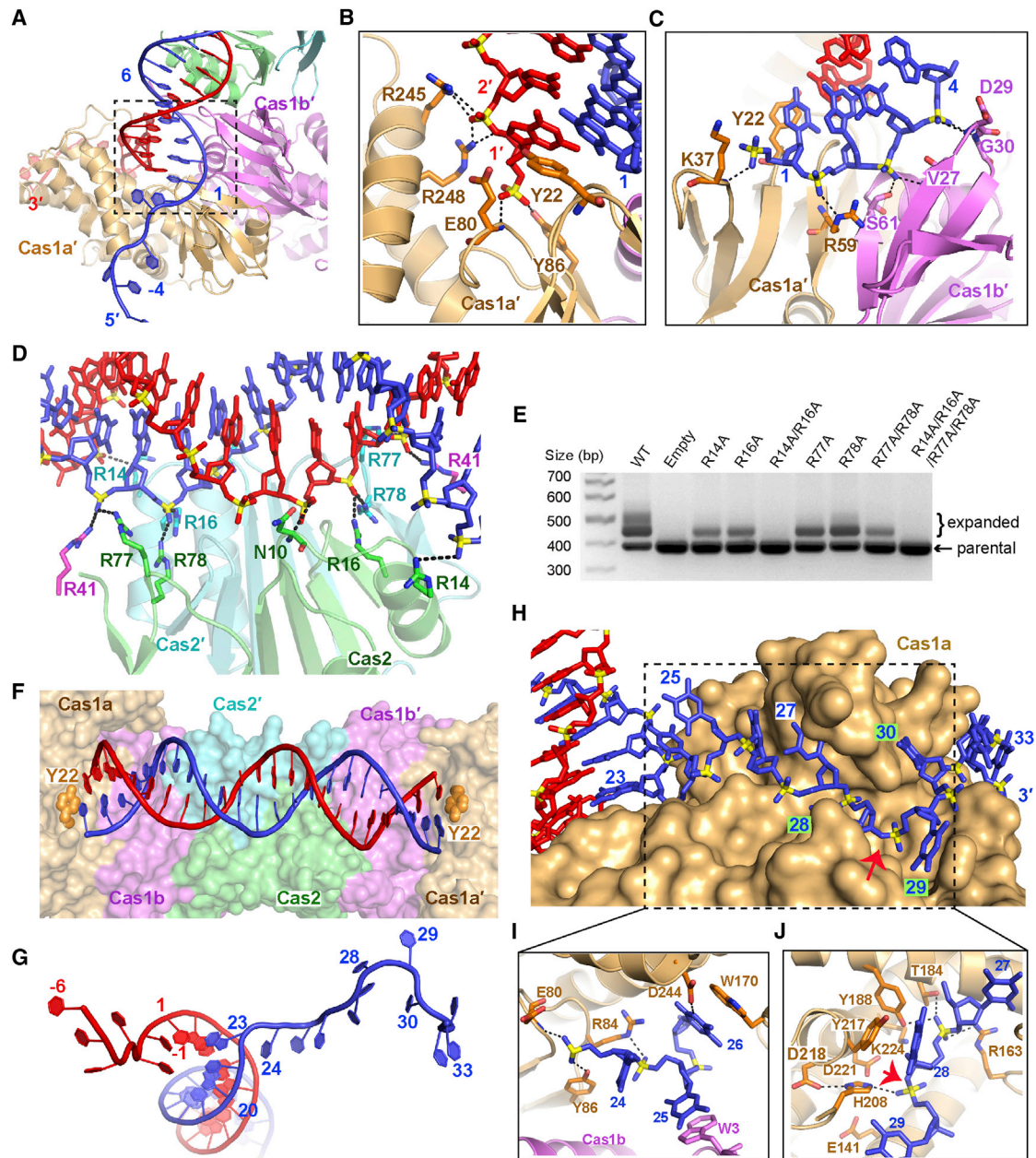
Cell

Figure 2. Positioning of Dual-Forked DNA onto Cas1-Cas2

(A) One terminus of the duplex straddles the Cas1 dimer interface.

(B and C) Detailed view of the interaction between Cas1 dimer and DNA duplex.

(D) Detailed view of interaction between Cas2 dimer and DNA duplex.

(E) Agarose gel of in vivo acquisition assays involving mutations of duplex-binding Cas2. WT, wild-type.

(F) Tyr22 residues from Cas1a and Cas1a′ bracket the 23-bp duplex, which is positioned on the flat surface of Cas1-Cas2.

(G) A simplified view (with Cas proteins removed) of the DNA 5′ and 3′ overhangs at one end of the complex.

(H) 3′ overhang lies in the groove of the C-terminal domain of Cas1a shown in surface view representation. The phosphate groups are shown in yellow. Nucleotides 28–30 are labeled with a green background, with the cleavage site shown by a red arrow.

(I) Magnified view of the interaction between nucleotides 24–26 and Cas1.

(J) Magnified view of the interaction between nucleotides 27–28 and Cas1. Glu141, His208, and Asp221 are the catalytic residues of Cas1. The DNA cleavage site is indicated by a red arrow.

See also Figure S2.

Cell

Two Tyrosine Residues Determine the 23 nt Length of the Bracketed Duplex Segment

Next, we investigated what specific interactions with the protein determine the length of the duplex segment in the complex. As shown in Figures 2B and 2F, the first base pair of the duplex stacks on the side chain of Tyr22 of Cas1a′, and the last base pair stacks on the Tyr22 of Cas1a. Such bracketing by the tyrosines prevents additional base pairs from participating in the duplex structure, with the tyrosines in addition serving as wedges that generate duplex single-strand junctions (Figure 2B). Thus, these two tyrosines from the symmetry-related Cas1a subunits serve as a caliper to measure a 23-bp duplex segment of the bound DNA (Figure 2F). In the case of this *E. coli* Cas1-Cas2-DNA complex structure, the distance between these two Tyr22 residues is 76 Å, creating a ruler that fits a B-form DNA duplex with the length of 22–23 bp.

To investigate whether the distance between the two Tyr22 residues is a function of the length of the duplex, we analyzed additional structures containing DNAs of shorter duplex length. Contrary to our expectations, the length of the duplex found in the structure of Cas1-Cas2 bound to the dual-forked DNA with 22-bp duplex was not 22 bp but, rather, 23 bp, identical to the complex containing 23-bp duplex dual-forked DNA discussed above (Figures S2C–S2E). Thus, the assembly of the Cas1 and Cas2 complex forms the basis for the two side chains of Tyr22 residues from Cas1a and Cas1a′ to work together as a ruler that defines the precise length of the duplex. In type I-E Cas1, Tyr22 is conserved to a certain extent, being always a planar/large side-chain residue (such as His or Arg), which could possibly also stack with the base pairs at both ends (Figure S3). Together, these observations strongly suggest that the duplex length is not simply a result of our DNA design but is a function of the intrinsic properties of the Cas1-Cas2-DNA complex. This explains how Cas1-Cas2 provides a ruler that measures with great precision the length of the DNA duplex.

3′ Overhangs Thread through the C-Terminal Domains of Cas1a

As the two Tyr22 residues act as wedges between the duplex and overhangs at the fork site, they cause a flip of the 3′ overhangs away from the duplex (Figure 2G). As a consequence, the 3′ overhangs thread through the C-terminal domain of Cas1a (Figure 2H) in a similar manner at both ends of the complex. The 10-nt 3′ overhangs (numbered 24–33) adopt an irregular curve-line conformation and form extensive intermolecular interactions with the C-terminal domains of Cas1a (Figure 2H). Nucleotide 24 flips away from the duplex, with its phosphate groups stabilized by hydrogen bonding to residues Glu80 and Tyr86 of Cas1a (Figure 2I). Nucleotide 25 is stabilized via stacking on the side chain of Trp3 of Cas1b, with further stabilization via interaction of its phosphate group with Arg84 of Cas1a (Figure 2I). Nucleotides 26–28 are stabilized via interactions with residues Trp170, Arg163, Thr184, Tyr188, His208, and Tyr217 of Cas1a (Figures 2I and 2J). Thus, these intermolecular interactions stabilize the bound single-stranded 3′ overhangs at either end, which is likely to be a pre-requisite for proper cleavage function of Cas1 (see below).

PAM Recognition

The molecular basis for the selection of the protospacer remains unknown. In *E. coli*, spacers are chosen from protospacer containing a 5′-AAG-3′ PAM sequence, and it was shown that the protospacer is cleaved between G-1 and A-2 within the PAM and that G-1 is inserted along with the protospacer (Datsenko et al., 2012; Goren et al., 2012; Swarts et al., 2012). In our structure, the cleavage is found between nucleotides 28 and 29 as described later, suggesting that nucleotides 28–30 in the 3′ overhang are complementary to the PAM sequence. Therefore, in vivo, these three nucleotides in the overhang should contain the sequence 5′-CTT-3′, as this is complementary to the PAM 5′-AAG-3′ sequence.

To provide insights into the molecular mechanism of PAM recognition by Cas1, we next determined the crystal structure of *E. coli* Cas1-Cas2 bound to DNA containing the PAM-complementary 5′-CTT-3′ sequence (Figures 3A–3C and Movie S1) instead of the original oligo-T sequence at positions 28–30. The overall structure of the PAM-complementary-containing complex is similar to the oligo-T-containing complex, though there are some important differences. Therefore, we will discuss below the PAM-complementary bound region, as well as those regions that differ between the PAM-complementary and oligo-T-bound structures of the complex. Given that the two 3′ overhangs bearing the PAM-complementary sequence insert into the C-terminal domain of Cas1a and Cas1a′ in the same manner, we will describe only the structural features of the 3′ overhang bound to Cas1a.

As shown in Figure 4A, seven nucleotides were visible at the 3′ overhang, where they adopt a hook-shaped curve and meander through the C-terminal domain of Cas1a. Nucleotides 24–27 are stabilized by Cas1a in the PAM-complementary-containing complex, in a manner similar to that observed in the oligo-T-containing complex described above. Nucleotides C28, T29, and T30 are positioned orthogonally to each of their preceding nucleotides and fit into a binding pocket provided by the C-terminal domain of Cas1a and the C-terminal tail of Cas1b. It is clear from the PAM-complementary-containing complex structure that this pocket is base specific for the CTT sequence. The nucleotide C28, which is complementary to the conserved G in the PAM sequence, is read out by two base-specific hydrogen-bonding interactions. The Watson-Crick edge of C28 forms a hydrogen bond with the side chain of Lys211 of Cas1a and with the non-bridging phosphate oxygen of nucleotide 27 (Figure 4B). The pyrimidine ring of C28 is further stabilized as a result of being sandwiched between the side chains of Tyr217 (Cas1a) and Ile291 (Cas1b) residues. The base of T29 is flexible in the oligo-T-containing structure. By contrast, in the PAM-complementary-containing complex, the base of T29 stacks on the side chain of Gln287 of Cas1b, with its Watson-Crick edge forming a base-specific hydrogen bond with the backbone oxygen of Arg138 of Cas1a. Further, the non-bridging phosphate oxygen atoms of T29 form hydrogen bonds with the side chains of His208 from Cas1a and Gln287 of Cas1b (Figure 4C). T30, whose base stacks on the side chain of Tyr165, is also recognized in a sequence-specific manner by forming hydrogen bonds involving its Watson-Crick edge with the main chain of Tyr165 in the PAM-complementary-containing complex (Figure 4C).
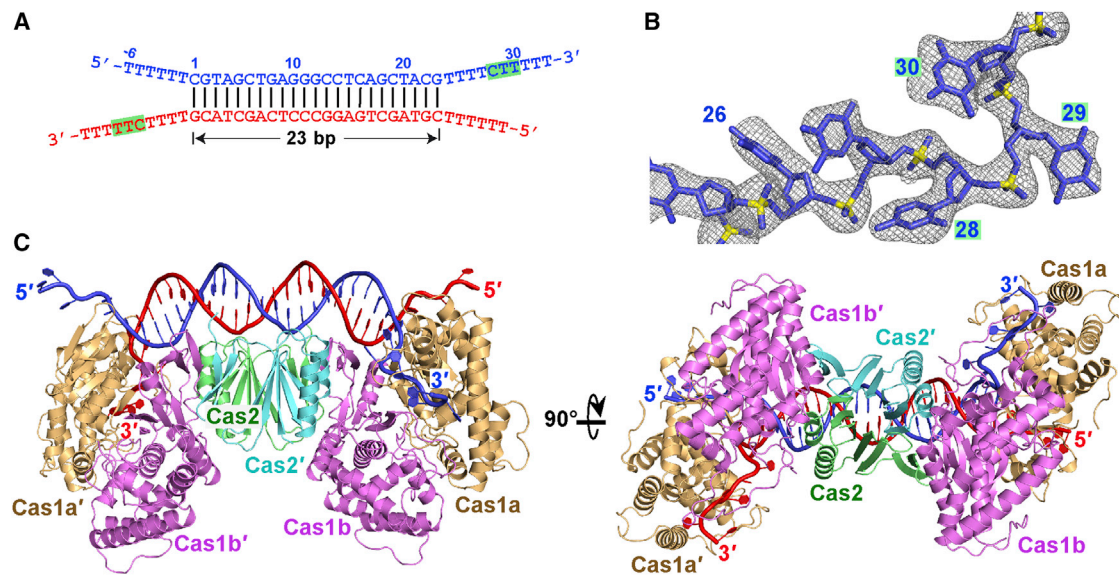
Cell

Figure 3. Crystal Structure of *E. coli* Cas1-Cas2 Bound to a PAM-Complementary Dual-Forked DNA

(A) Schematic diagram of the PAM-complementary dual-forked DNA, which is a 23-mer palindromic duplex with $5'$-$(T)_6$ and $3'$-$(T)_{10}$ overhangs on both ends. The PAM-complementary sequence $5'$-CTT-$3'$ is highlighted by the green background.

(B) Fo-Fc omit map (gray color, contoured at 3.0 σ) of the nucleotides 26–30 in the structure with the PAM-complementary sequence within the $3'$ overhangs.

(C) Orthogonal views of the crystal structure of the complex of Cas1-Cas2 bound to the PAM-complementary dual-forked DNA.

See also Figure S3 and Movie S1.

To investigate how the base-specific interaction between Lys211 and C28 is related to conservation of the G residue, which is present at the $5'$ end of most of the newly acquired spacers, we sequenced newly acquired spacers within either wild-type or the Lys211Ala Cas1 mutant. We found that, in the wild-type Cas1, ~76% new spacers are flanked by a $5'$ G, whereas it is reduced to 47% in the Lys211Ala mutant. The Watson-Crick edge of C28 is recognized in a sequence-specific manner via two hydrogen bonds. Removing one base-specific interaction with C28 by substituting Lys211 with Ala markedly decreased the degree of G conservation (Figure 4D). Thus, the interaction between the bases of C28 and Lys211 is important for the insertion of the conserved G.

Single-Stranded Nature of the $3'$ Overhang Is Critical for New Spacer Acquisition

To test the significance of the $3'$-terminal single strand and the PAM-complementary sequence, we conducted electrophoretic mobility shift assays (EMSA). As shown in Figure 4E, the presence of $3'$ overhangs significantly increases the binding affinity between Cas1-Cas2 and DNA. Cas1-Cas2 binds blunt-end double-stranded DNA with lower affinity than dual-forked DNA. Using a DNA duplex flanked by a 4-nt $3'$ overhang at both ends moderately increased the affinity for Cas1-Cas2. However, the binding affinity increased significantly upon extension of the $3'$ overhang by either 7 or 10 nt, with no further change on proceeding from 7–10 nt. By contrast, weak binding was observed when the DNA substrate contained 10-nt $5'$ overhangs (Figure S4A), implying a modest contribution to binding from the $5'$ overhang. Most importantly, the binding is much stronger when the 7-nt $3'$

overhang contains the PAM-complementary $5'$-CTT-$3'$ sequence (Figure 4E) compared to $5'$-TTT-$3'$, $5'$-TCC-$3'$, and $5'$-GAA-$3'$ sequences (Figure 4F), establishing that $5'$-CTT-$3'$ of the PAM-complementary sequence is crucial for high-affinity protospacer binding by Cas1-Cas2.

Impact of DNA-Binding Cas1 and Cas2 Mutants on Complex Formation

As shown in Figure 4A, the $3'$ overhangs are located within the C-terminal domain of Cas1a, where they are stabilized by numerous intermolecular interactions (Figure 5A). With the exception for the PAM-complementary sequence, the $3'$ overhangs bind to the Cas1 dimer mainly through non-sequence-specific interactions. Aromatic residues Tyr165, Trp170, and Tyr 217 on Cas1a are involved in stacking interactions with the bases of the $3'$-overhang segment. We observe in an EMSA assay a modest decrease in binding affinity for the alanine-substituted Tyr165 and Trp170 dual mutant, while a more pronounced decrease is observed for the Tyr165 and Tyr217 dual mutant (Figure 5B, top), with the latter two involved in complementary-PAM recognition (Figures 4B and 4C). In addition, a significant reduction in binding affinity is observed for alanine-substituted Arg14 and Arg16 dual mutant (Figure 5B, bottom), consistent with these Cas2 residues involved in intermolecular recognition with the duplex segment (Figure 5A).

Impact of DNA-Binding Cas1 Mutants on In Vivo Spacer Acquisition

Tyrosine residues 165, 188, and 217, as well as Lys211 on Cas1a, are involved in intermolecular recognition of the

Cell
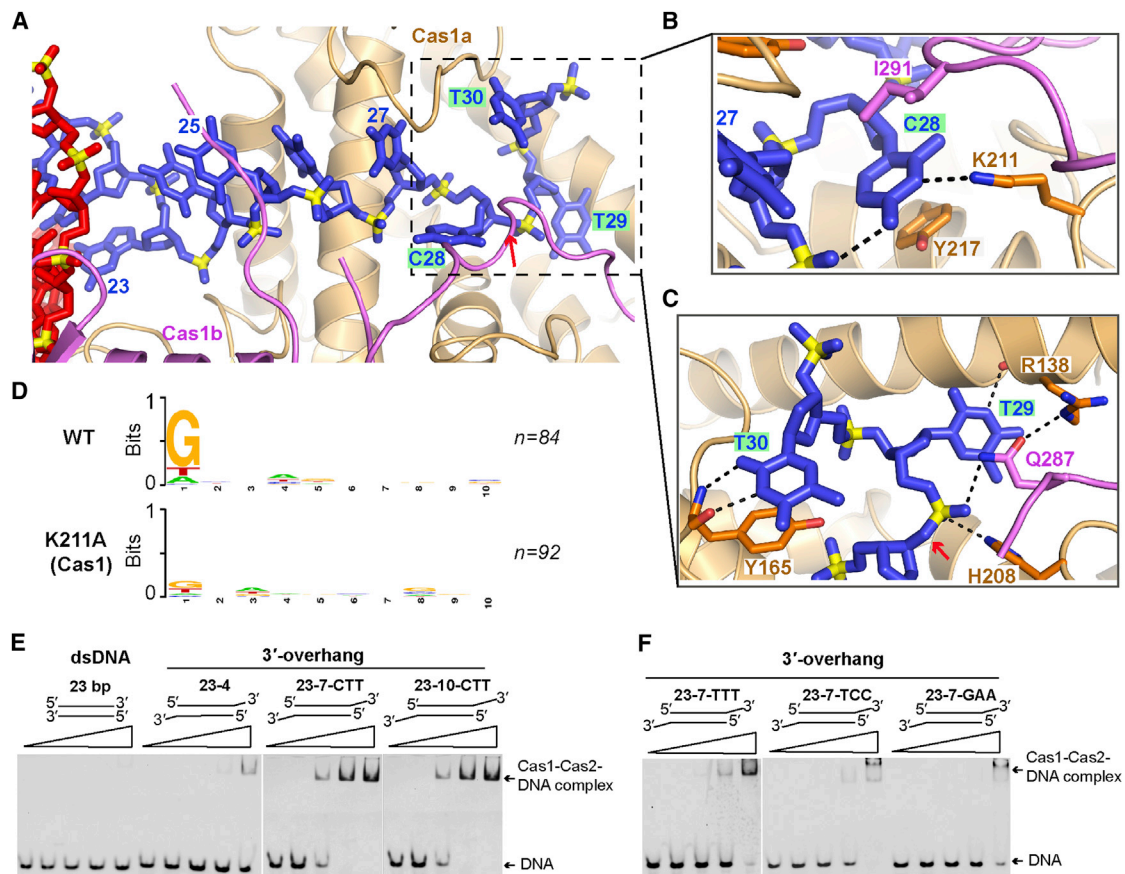
Figure 4. PAM-Complementary Segment Recognition

(A) The 3′ overhang containing the PAM-complementary sequence motif lies in the groove of the C-terminal segment of Cas1a and covered by the C-terminal tail of Cas1b. The nucleotides complementary to the PAM are labeled by green background.

(B and C) The detailed sequence-specific interactions between Cas1 and C28 (B) and T29 and T30 (C) residues. The DNA cleavage site is indicated by a red arrow in A and C.

(D) Sequence logos obtained after the alignment of the first ten nucleotides of the new insertion. Numbers indicate the positions of the nucleotide of the new insertion. Number of sequences used in each alignment is indicated as *n*.

(E) Electrophoretic mobility shift assay using 5′ Cy3-labeled double-stranded DNA-containing 23-bp duplex and the 23-bp duplex with 4-, 7-, or 10-nt 3′ overhangs on both ends. The 23-7-CTT and 23-10-CTT DNAs harbor the PAM-complementary sequence 5′-CTT-3′, as shown in Table S2.

(F) Electrophoretic mobility shift assay using 5′ Cy3-labeled non-PAM-complementary DNAs with 23-bp duplex and 7-nt 3′ overhangs. The PAM-complementary sequence 5′-CTT-3′ was replaced by 5′-TCC-3′, 5′-GAA-3′, or 5′-TTT-3′.

See also Figure S4.

PAM-complementary sequence of the 3′ overhang in the Cas1-Cas2-DNA complex (Figures 4B, 4C, and 5A). Replacement of individual Tyr165, Tyr188, and Tyr217 by alanine resulted in significant reduction in spacer acquisition in an in vivo assay, while a modest reduction was observed for the Lys211Ala mutant, as shown in Figure 5C, top. Interestingly, Tyr22, which is involved in bracketing the duplex segment (Figure 2F), shows only a modest decrease in spacer acquisition on replacement by alanine (Figure 5C, top). This was unanticipated but may reflect the dominant role of intermolecular interactions involving the 3′-overhang segment to generation of the duplex single-strand junction, as reflected in loss of spacer acquisition for the Arg245Ala and Arg248Ala dual mutant (Figure 5C, bottom) that is positional at the junctional site (Figure 5A).

Identification of the Cleavage Site within the 3′-Overhang Segments

The nuclease activity of Cas1 is crucial for new spacer acquisition, with conserved residues His208, Glu141, Asp221, and Asp218 crucial for this function (Nuñez et al., 2014). In our structure of the complex, the phosphate group of nucleotide 29 is positioned adjacent to the side chains of His208, Glu141, and Asp221 that line the catalytic pocket, with the side chain of His208 forming a hydrogen bond with the phosphate group of T29 (Figure 5D). This suggests that Cas1 cleaves the phosphodiester bond between nucleotides 28 and 29, resulting in a DNA cleavage product that contains a 5-nt 3′ overhang (Figure 5E). We thus performed a cleavage assay using a 23-bp duplex DNA flanked by 10-nt 3′ overhangs at either end. As shown in Figure S4B, the cleavage product is indeed 5 nt shorter,

Cell
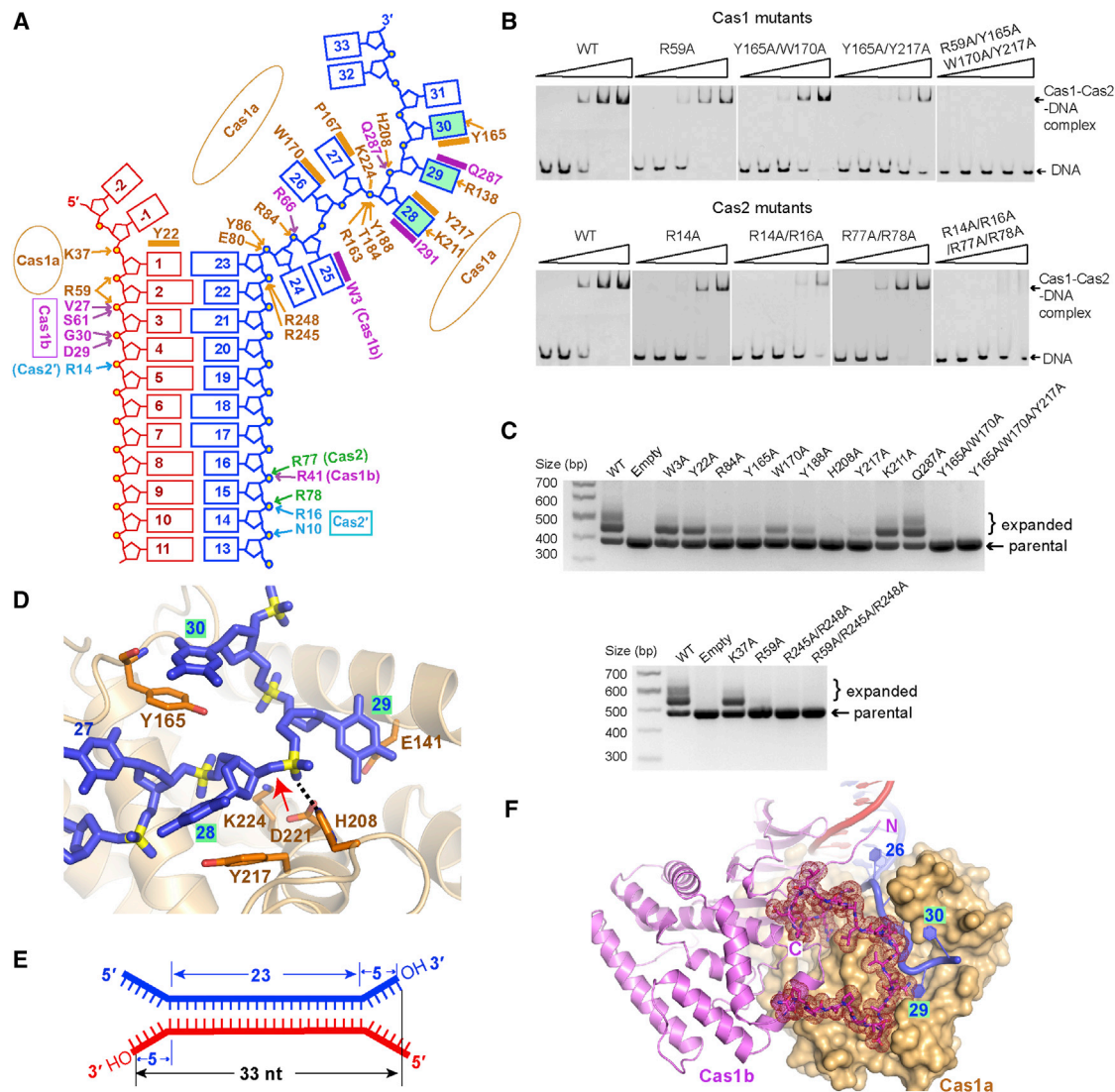
Figure 5. The C-Terminal Domain of Cas1a Recognizes the PAM-Complementary Sequence

(A) A schematic listing intermolecular contacts in the crystal structure of Cas1-Cas2 bound to a palindromic dual-forked DNA.

(B) Electrophoretic mobility shift assay using 5′ Cy3-labeled 23-bp duplex with 7-nt 3′ overhangs (DNA 23-7-CTT), involving mutations of Cas1 (top) or Cas2 (bottom).

(C) Agarose gels of in vivo acquisition assays involving mutations of Cas1.

(D) Zoomed-in view of the catalytic site with nucleotides 28 and 29 located in the catalytic pocket. The DNA cleavage site is highlighted by a red arrow.

(E) Schematic diagram of Cas1 cleavage product.

(F) The C-terminal tail of Cas1b, which is shown in stick representation and magenta mesh density, covers the catalytic pocket of Cas1a.

See also Figure S5.

confirming the proposed cleavage site. Here, seven nucleotides within the 3′-terminal overhangs are observed in our structure, suggesting that Cas1-Cas2 binds an intact substrate. Another residue, Asp218, was previously thought to be a catalytic residue (Babu et al., 2011). However, in our structure, it is positioned away from the catalytic pocket and does not directly contact the DNA substrate. Instead, it stabilizes the alignment of the conserved catalytic residue His208 via a hydrogen bond (Figure 2J).

PAM-Complementary Sequence Stabilizes C-Terminal Tail of Cas1b

We compared the structures of PAM-complementary-containing complex and oligo-T-containing complex to highlight the conformational change upon binding of the PAM-complementary sequence. As shown in Figure S5A, the overall structures of these two complexes are similar, though there are distinct differences. Thus, in the complex containing oligo-T DNA, the proline-rich C-terminal tails of Cas1b and Cas1b′ are disordered. By

Cell

contrast, in the PAM-complementary-containing complex, the C-terminal tails of Cas1b and Cas1b′ are well ordered and are involved in the binding of the PAM-complementary sequence (Figures 4B and 4C). In the PAM-complementary-containing complex, the loop containing the residues 278–305 of Cas1b covers the catalytic pocket of Cas1a, similar to a lid-like topology (Figures 5F and S5B). Residues Ile291 and Gln287 in the C-terminal tail are involved in the interaction with the PAM-complementary sequence (Figures 4B and 4C), suggesting that the interactions between the PAM-complementary sequence and the C-terminal tail of Cas1b stabilize the fold of the latter. Interestingly, in the DNA-free complex (PDB: 4P6I), the C-terminal tail of Cas1b is ordered and spans Cas2 (Figure S5C) (Nuñez et al., 2014). In the PAM-complementary-containing complex, the C-terminal tail of Cas1b does not span Cas2 any longer but covers the catalytic pocket of Cas1a (Figure S5B).

The Conformational Changes of Cas1-Cas2 upon Protospacer Binding

To investigate whether the binding of the protospacer causes structural rearrangements of Cas1-Cas2, we performed comparative superposition analysis. Comparison of the DNA-free (Figure S6A) and DNA-bound (Figure S6B) structures reveals that the protospacer binding triggers large structural rearrangements in Cas1-Cas2. The Cas1-Cas2 in its DNA-free state adopts a "wings-up" butterfly-shaped configuration, in which the four Cas1 monomers represent the wings and the Cas2 dimer represents the body (Figure S6C, left). Superposition of the Cas2 dimer of the free and DNA-bound structures shows that the two Cas1 dimers rotate in either clockwise (Cas1a/b) or anti-clockwise (Cas1a′/b′) directions upon complex formation (Figures 6A, S6A, and S6B), similar to butterfly wings dropping into a spread-out position (Figure S6C, right). This conformational change of the Cas1-Cas2 likely facilitates new spacer incorporation into the CRISPR locus. First, this rotation results in the generation of a flat protein surface for binding the duplex segment of the bound DNA (Figure 1D). Second, this rotation repositions the two tyrosine residues from Cas1a and 1a′ into forming a bracket that precisely spans the full duplex length (Figure S6D). Third, the rotation and loop (residues 163–174 of Cas1a) movement results in the formation of an optimal catalytic pocket within Cas1a, allowing site-specific cleavage (28–29 step) within the 3′ overhang (Figure 6B). Fourth, it creates a deep arch-shaped surface on the opposite face of the duplex-binding surface (Figure 1D).

To understand what induces the conformational change of Cas1-Cas2 upon protospacer binding, we superimposed either Cas2 or Cas1b′ in their DNA-free and DNA-bound states (Figures 6C and 6D). As shown in Figure S6E, two antiparallel β strands (β6-β7) of Cas2 interact with Cas1b. A comparison of Cas2 structures in the DNA-bound and DNA-free Cas1-Cas2 (PDB: 4P6I) shows that β6-β7 of Cas2 undergoes a significant conformational change (Figure 6C). Upon protospacer binding, Arg77 of Cas2, which is positioned in the loop linking β6-β5, flips by 180 degrees, allowing formation of an interaction with the DNA duplex (Figure 6C). The downstream residue Arg78 is also involved in duplex DNA binding (Figure 2D). Together, as a consequence of these interactions, the β6-β7 sheet moves

away (see yellow arrow, Figure 6C) from the core ferredoxin fold of Cas2.

Next, we compared the structures of the Cas1-Cas2 interface by superimposing Cas1b′ within the DNA-free and DNA-bound complexes. With Cas1b′ well superposed, Cas2 and Cas1a/b rotate away from the DNA-binding interface, as indicated by the yellow arrow (Figures 6D, S6F, and S6G). Interestingly, β6-β7 of Cas2 also superposed well along with Cas1b′ during this superimposing of free and bound states (Figure 6D), suggesting that the binding of the protospacer does not affect Cas1-Cas2 interaction and that the loop linking β6 and the core ferredoxin fold of Cas2 plays an essential role in the hinge-mediated movement upon protospacer binding.

DISCUSSION

In this structural study, we reveal the precise nature of the DNA substrate of Cas1-Cas2. Furthermore, we provide evidence that the structural properties of this complex are the basis for the strict length requirements observed for newly acquired spacers incorporated into the CRISPR array. Lastly, we identify the mechanisms behind the selection of the protospacer sequence, namely by Cas1-Cas2 recognizing the PAM-complementary sequence in the invading DNA.

Cas1a and Cas1b Subunits Perform Different Functions during Acquisition

Cas1 proteins are asymmetrical homodimers, whereby two Cas1 monomers forming the dimer adopt different conformations, in relation to the relative orientations between the N- and C-terminal domains (Figure 6E). The asymmetry of the Cas1 dimer was also observed in DNA-free *E. coli* Cas1-Cas2 (Nuñez et al., 2014) and in DNA-free Cas1 dimers from other organisms (Babu et al., 2011; Kim et al., 2013; Wiedenheft et al., 2009). This indicates that it is a common feature of Cas1 that its two monomers within the dimer adopt different conformations, which implies that these two monomers are likely to have different biological functions.

As shown in Figure 4A, the 3′ overhang inserts into the C-terminal domain of Cas1a and threads through the catalytic site. The 5′ overhangs interact with the C-terminal domain of Cas1b or Cas1b′ that belong to two neighboring symmetric complexes in the crystal lattice (Figure S6H). However, it is unclear whether this latter structural feature results from complex formation or from crystallographic packing of another complex next to the 5′ overhangs. Nevertheless, the possibility can be excluded that the 5′ overhangs bind to Cas1b or Cas1b′. In our structures, Cas1b and Cas1b′ form contacts on either side of the Cas2 dimer, while no contacts are observed between Cas1a or Cas1a′ with the Cas2 dimer. Arg245 and Arg248 in Cas1b are involved in interaction with Cas2′ (Figure S6E), whereas these residues in Cas1a interact with the DNA duplex (Figure 2B). Together, each asymmetrical Cas1 homodimer possesses one catalytic subunit (Cas1a and Cas1a′), which generates a 3′-OH group following cleavage and for recognition of the PAM-complementary sequence to select the protospacer, and one subunit (Cas1b and Cas1b′), which is responsible for forming Cas1-Cas2. Thus, our structure sheds light
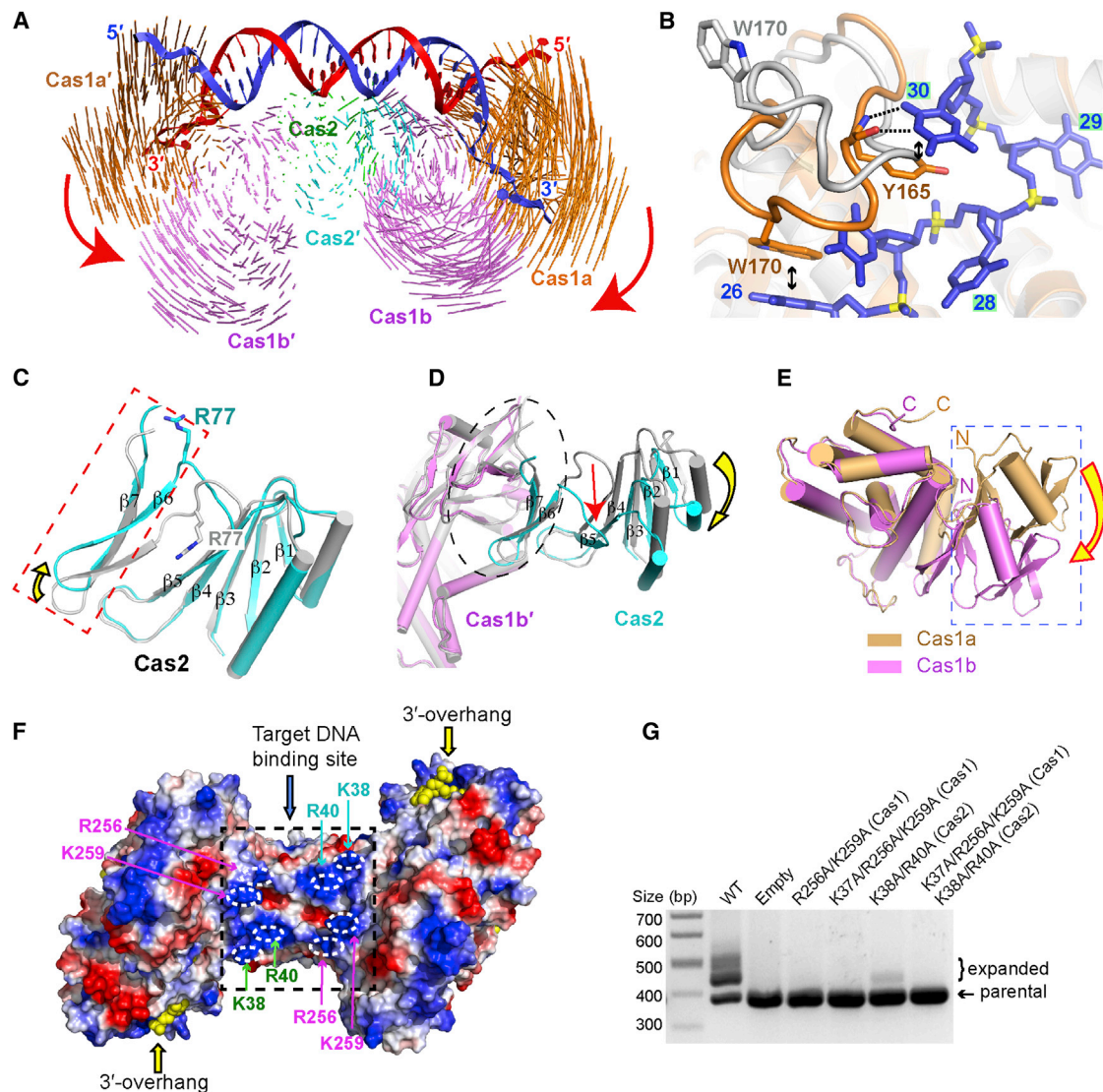
Cell

Figure 6. Conformational Change of Cas1-Cas2 upon Formation of Protospacer-Bound State and Function of Cas1 and Cas2 Proteins

(A) Structural comparison between Cas1-Cas2 in the protospacer-bound and DNA-free (PDB: 4P6I) structures. The Cas2 protein is superimposed. Vector length correlates with the domain motion scale. The red arrows indicate domain movements within Cas1-Cas2 complex upon protospacer binding.

(B) The loop from residues 163 to 174 adjacent to the catalytic pocket undergoes a conformational change upon binding of the 3′ overhang bearing PAM-complementary sequence (note the shift from silver to orange representations). The stacking interactions are highlighted by black double-edged arrows.

(C) Structural comparison of Cas2 in the protospacer-bound Cas1-Cas2 complex (in cyan) and DNA-free Cas1-Cas2 structure (in silver). There is good superposition for the core ferredoxin fold of Cas2. The yellow arrow indicates the movement of β6-β7 of Cas2. Residue R77, which undergoes a significant conformational change, is shown in a stick representation.

(D) Structural comparison of Cas1b in the protospacer-bound Cas1-Cas2 complex and DNA-free Cas1-Cas2 structure. There is good superposition for Cas1b and β6-β7 of Cas2. The yellow arrow indicates the movement of the core fold of Cas2. The red arrow indicates the movement of the loop linking β6 and the core fold of Cas2.

(E) Superposition of the catalytic domain of Cas1a (light orange) and Cas1b (magenta). The yellow arrow shows the conformational difference of the N-terminal domain.

(F) The arch-like structure may involve a binding site for the target DNA within its positive charged patches highlighted by a black box. The Cas1-Cas2 complex is shown as a surface representation and is labeled according to its electrostatic potential (red, negative charge; blue, positive charge). The DNA is shown as yellow spheres.

(G) In vivo acquisition assay with potential Cas1 and Cas2 mutations positioned within the postulated target DNA binding sites.

See also Figure S6.

Cell

on the question of why Cas1 dimers are asymmetric, with the subunits fulfilling two different functions.

Function of Cas2

Our structures of the complexes also shed light on the role of Cas2 during CRISPR adaption. The Cas2 dimer bridges two Cas1 dimers, forming Cas1-Cas2, which then provides the binding surface for the protospacer DNA. Together with two Cas1 dimers, Cas2, acting as a space holder, measures the length of the duplex by ensuring that the Tyr22 residues of Cas1a and Cas1a′ are positioned exactly 23 nt apart from each other (Figure 2F). Moreover, the Cas2 dimer plays crucial roles in stabilizing the bound duplex DNA by forming hydrogen bonds with the backbone of the DNA duplex (Figure 2D). Also, opposite to the duplex binding surface of Cas2 is an arch-like structure, which is likely to be involved in recognition of the target DNA, based on our observation that the arch topology contains positively charged patches formed by residues Lys38 and Arg40 of Cas2 and Arg256 and Lys259 of Cas1b (Figure 6F). Notably, Lys38Ala and Arg40Ala (Cas2) dual mutant significantly reduced spacer acquisition, while no insertion was observed for the Arg256Ala and Lys259Ala (Cas1) dual mutant (Figure 6G). However, further studies will be required to verify the target DNA binding site. Thus, the Cas2 dimer acts as an adaptor protein, bringing two Cas1 dimers together while stabilizing and measuring the length of the protospacer DNA, as well as binding to the target DNA.

Cas1-Cas2 Predetermines the Length of the Protospacer

Our structural analysis revealed that the most promising substrate of Cas1-Cas2 is composed of a dual-forked DNA, which contains both a double-stranded duplex and 3′ single-stranded overhangs on both ends. Importantly, the site of interaction involving the catalytic residues with the DNA is 5 nt away from the end of the duplex (Figure 5D). Thus, the putative DNA fragment contains 23 nt of the duplex region, as well as 5-nt 3′ overhangs at both ends, resulting in a total distance of 33 nt from one cleaved 3′ end to the other (Figure 5E). This finding is consistent with a recently proposed model (Nuñez et al., 2015), which suggests that Cas1-Cas2 inserts the invading DNA into the CRISPR locus like an integrase, with the length of the newly acquired spacer in the CRISPR locus depending on the 3′ ends of the two strands of the protospacer DNA. Therefore, our structures of the Cas1-Cas2-DNA complex most likely represent the Cas1-Cas2-protospacer-containing DNA complex. These structures provide insights into how Cas1-Cas2 predetermines the length of protospacer by utilizing two Tyr22 residues to measure a 23-bp duplex, and the positioning of the catalytic residues determines the cleavage position, thereby generating 5-nt 3′ overhangs on both strands. Thus, the architecture of the Cas1-Cas2-protospacer DNA complex provides the basis for the observed length of 33 nt of the DNA cleavage product, thereby explaining what factors contribute to the determination of the constant length of newly acquired spacer in vivo.

Source of Protospacer

Prior to our study, the exact nature of the DNA substrate associated with Cas1-Cas2 was unknown. Here, we reveal that, apart from a double-stranded duplex region, single-stranded overhangs are critical for DNA-protein complex formation. We show that the unique interaction between the 3′ overhang and the catalytic domain of Cas1a is possible for ssDNA overhangs, but not for rigid dsDNA duplexes. In addition, our binding assay suggests that a 3′ overhang containing a minimum of 7 nt is essential for the association between Cas1-Cas2 and the DNA substrate (Figure 4E), possibly because the last 3 nt (positions 5–7 in the overhang) are, in fact, complementary to the AAG-containing PAM sequence in the invading DNA, thereby explaining why each new spacer starts with a G residue (Yosef et al., 2013).

If our model is correct, the question arises as to where such a single-stranded protospacer overhang would occur in an in vivo situation, i.e., in the invading phage or plasmid DNA. Intriguingly, spacer acquisition was shown to be highly replication dependent. The DNA degradation intermediates of RecBCD complex present at stalled replication forks might be the source of new spacers, as these intermediates include ssDNA fragments and degraded dsDNA (Levy et al., 2015; Paez-Espino et al., 2013). This finding fits well with our analysis and addresses the question of the origin of the single-stranded protospacer 3′ overhang (Figure 7A), and our results also address why the protospacer hotspots are located between sites of stalled replication forks and Chi sites. Together, our structures strongly suggest that E. coli protospacers are recognized and associated with Cas1-Cas2 in a dual-forked DNA topology, consisting of a 23-bp duplex and a minimal 7-nt single-stranded 3′ overhang in vivo. Therefore, in addition to the PAM that affects the spacer choice, the structural feature of the protospacer DNA also influences the frequency of protospacer incorporation.

Protospacer Selection

The interactions observed in our structure between Cas1a and 5′-CTT-3′ (Figures 4B and 4C), together with the EMSA results indicating the minimal 7-nt length requirement of the 3′ overhangs (Figure 4E), strongly suggest that the PAM-complementary sequence (being the last three nucleotides in the 7-nt 3′ overhang) plays a significant role in ensuring proper complex formation. In agreement with the important role of the length of 3′ overhangs, the complex of Cas1-Cas2 co-crystallized with single-forked DNA containing 10-bp duplex and only six T overhangs at both 3′ and 5′ ends was free of DNA, indicating insufficient association between DNA and protein complex. Together, these findings support the notion that 3′ overhangs of defined length and the PAM-complementary sequence are both essential for DNA binding to Cas1-Cas2 and thus critical for spacer acquisition. In all likelihood, these results explain the observation that AAG motif in the PAM sequence enhances adaption of the protospacer adjacent to it (Yosef et al., 2013).

PAM recognition is essential for protospacer selection during acquisition and for target selection during crRNA interference (Deveau et al., 2008; Mojica et al., 2009). In the acquisition machinery of E. coli type I system, Cas1a recognizes the PAM-complementary sequence in its single-stranded form. In the type II system, during crRNA interference, the target DNA flanked by PAM sequence (5′-NGG-3′) is recognized by Streptococcus pyogenes Cas9 in its dsDNA form (Anders et al., 2014). Interestingly, in the type II CRISPR-Cas system, Cas9 is not only
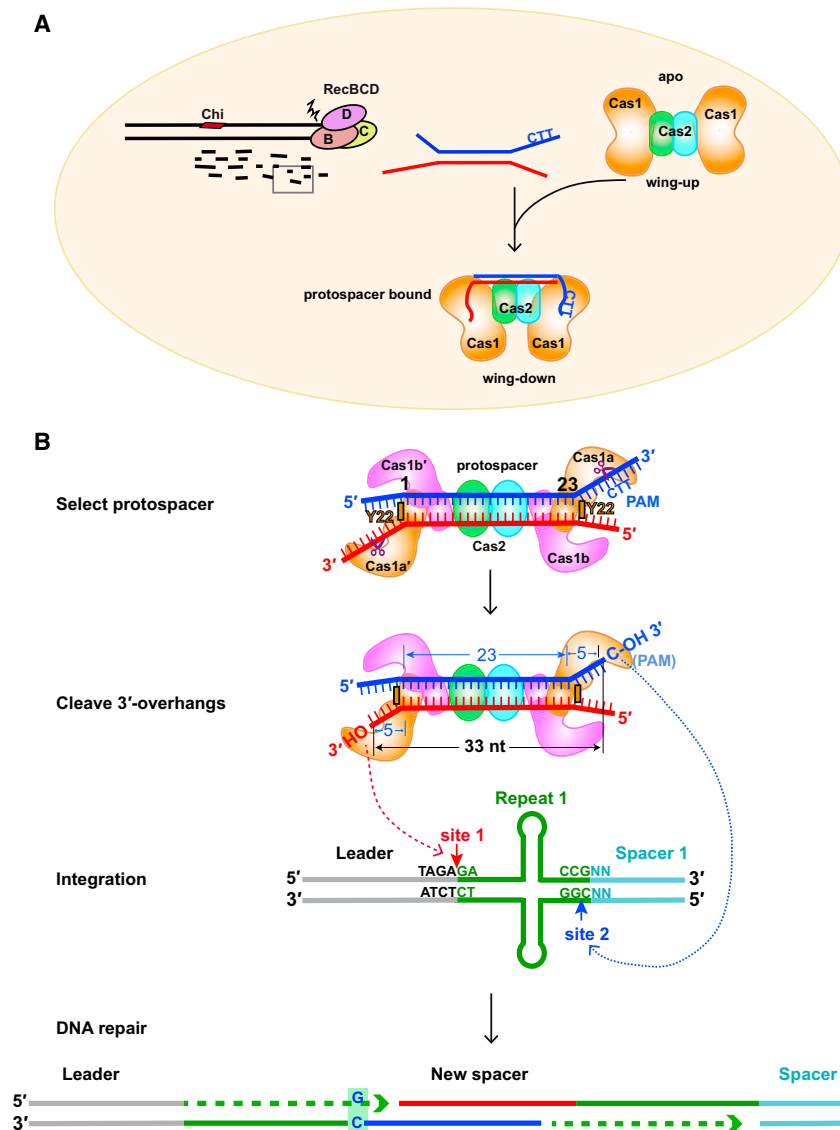
Cell

Figure 7. Model of CRISPR Spacer Acquisition

(A) Model explaining the capture of new DNA sequences from invading nucleic acid. Note the schematic representations of the "wing-up" and "wing-down" conformations of the apo- and protospacer-bound Cas1-Cas2 complexes. To simplify, both monomers in a Cas1 dimer are in orange.

(B) Model of DNA integration into the host CRISPR array. The Cas1a-mediated cleavage sites located on the 3′ overhangs, which are positioned 5 residues from the terminal base pairs, are represented by purple scissors. The cleavage product has 5-nt 3′ overhangs with 3′-OH groups on both strands, resulting in a distance between the 3′ overhang ends of 33 nucleotides. The two 3′ ends of the incoming protospacer are involved in nucleophilic attack on the CRISPR locus, as shown by the dashed red and blue arrows, respectively. Lastly, the gapped duplex is repaired by the host DNA replication machinery. The GC base pair originated from the PAM sequence is highlighted by green background. The leader is in gray, repeat 1 in green, and spacer 1 in cyan.

age, Cas1-Cas2 catalyzes the integration of the incoming DNA at the leader end of the CRISPR locus by two nucleophilic attacks at two sites on opposing strands (Nuñez et al., 2015; Rollie et al., 2015). The leader-Repeat1 segment is asymmetric, and the two sites on the target DNA have different sequences, with the choice of 3′-OH selection based on the terminal residue being a C. As shown in Figure 7B, site 2 (5′-CGG-3′) may preferentially select the 3′-OH of C. Thus, the leader sequence and the sequences surrounding the protospacer integration sites may play a critical role in correctly orienting the 3′-OH of C end of the protospacer DNA substrates for incorporation within the CRISPR locus. A recent study found that an artificial leader-Cas combination results in the insertion of the complex in the wrong orientation (Díez-Villaseñor et al., 2013). This observation is consistent with our model shown in Figure 7B. Therefore, we speculate that the sequence of leader-repeat 1 within the CRISPR locus may affect the binding orientation of the Cas1-Cas2-protospacer complex on the CRISPR locus.

CRISPR Adaption Likely Works through a Cut-and-Paste Mechanism

As Cas1 and Cas2 proteins are essential in both naive and primed adaptation, we propose that our structures of the complexes are likely to be suitable for both types of immunity. During primed adaptation, the partial ssDNA, resulting from the Cas3 degradation product or from an R loop formed

involved in the interference, but also in the spacer acquisition by associating with Cas1, Cas2, and Csn2 forming the acquisition machinery, thus coupling the interference and the acquisition machineries (Heler et al., 2015).

In the spacer acquisition step of both type I and II systems, Cas1 and Cas2 are critical, and the cleavage activity of Cas1 is essential for acquisition. By contrast, Cas9 binds the PAM in the type II system, while Cas1 recognizes the PAM-complementary sequence in the type I system. Whether Cas1 is also involved in the protospacer selection in the type II system remains under debate.

Mechanism of CRISPR Acquisition

Given that Cas1-Cas2 is symmetric, both Cas1a and Cas1a′ are capable of recognizing and binding the PAM-complementary sequence (5′-CTT-3′) and cleaving the overhangs of the protospacer to generate two 3′-OH groups. Following cleav-

Cell

upon crRNA binding target DNA, might be used as a precursor for new spacers by Cas1-Cas2. To date, the general assumption is that spacer acquisition works through a copy-and-paste mechanism, as opposed to a cut-and-paste process. Our structures reveal that Cas1 selects and cuts the foreign DNA to generate a spacer, which is in agreement with previous studies stating that Cas1-Cas2 mediates the cleavage-ligation reaction (Arslan et al., 2014), indicating that the CRISPR adaption likely works via a cut-and-paste mechanism.

The acquisition of new spacer sequences is absolutely essential for acquiring immunological memory and is crucial for maintaining an advantage over invading DNA elements by continuously updating the DNA library for crRNA interference of invading DNA elements. Our study shows that *E. coli* possesses a sophisticated machinery that utilizes frequently occurring PAM sequences as essential identification markers, which allow for efficient cleavage of the DNA sequence once embedded into Cas1-Cas2. Therefore, Cas1-Cas2 acts as a sequence-specific integrase. Of equal importance, this protein complex was designed by nature in such a manner that the protospacer binding results in a major conformational change in the protein, in the process of which an arch-like structure is created that is likely to be involved in proper binding to the first repeat of the CRISPR locus. These findings should lay the foundation and greatly facilitate the quest for identifying additional insights into the structural mechanisms responsible for the integration of new spacers into the CRISPR locus.

EXPERIMENTAL PROCEDURES

Detailed experimental procedures are described in the Supplemental Experimental Procedures.

E. coli Cas1 and Cas2 were cloned into pET-sumo expression vector and expressed in *E. coli* Rosetta2 (DE3) (Novagen). Cas1 was purified by chromatography on nickel and Heparin HP column (GE Healthcare). Cas2 was purified by chromatography on nickel, Q FF column, and Superdex 200 (GE Healthcare). Cas1 and Cas2 proteins were concentrated to 35 mg/ml and 5 mg/ml, respectively. The Cas1 and Cas2 mutants were made with Quick-Change kit and verified by sequencing. All mutant proteins were expressed with the same protocol as that used for the wild-type protein.

The Cas1-Cas2 single-forked DNA complex was reconstituted by incubating Cas1, Cas2, and single-forked DNA at the molar ratio of 1:1.1:0.6 on ice for 30 min and was further purified by gel filtration. The Cas1-Cas2 dual-forked DNA complex was reconstituted on ice for 30 min by incubating Cas1, Cas2, and DNA at the molar ratio of 1:1.1:0.3.

The Cas1-Cas2-DNA complexes were crystallized at 16°C by the hanging-drop vapor diffusion method. All Cas1-Cas2-DNA complex crystals were obtained by mixing equal volumes of complex solution and reservoir solution. X-ray diffraction data were collected at 100 K on the beamlines BL-17U and BL-19U at Shanghai Synchrotron Radiation Facility. All structures were solved by molecular replacement using the Cas1 monomer and Cas2 monomer in the DNA-free Cas1-Cas2 structure as the search models. All structures were refined using the program Refmac and Phenix and were manually built with COOT. All structural figures were prepared with Pymol (http://pymol.org).

Binding affinities of various DNA molecules to Cas1-Cas2 were tested using an EMSA. Functional importance of DNA-interacting residues was validated by EMSA and by using an in vivo spacer acquisition assay, as described previously (Yosef et al., 2012). Furthermore, the cleavage assays were undertaken using 5′ Cy3-labeled DNA with 23-bp duplex flanked by 10-nt 3′ overhangs. The sequences of all DNA oligonucleotides used in the study are listed in Table S2.

ACCESSION NUMBERS

The atomic coordinates of the Cas1-Cas2-DNA complexes have been deposited in the Protein Data Bank with accession numbers listed in parenthesis. Cas1-Cas2 single-forked DNA (PDB: 5DQU), Cas1-Cas2 dual-forked DNA with 23-bp duplex (PDB: 5DLJ), Cas1-Cas2 dual-forked DNA with 22-bp duplex (PDB: 5DQT), and Cas1-Cas2 bound to the PAM-complementary sequence (PDB: 5DQZ).

SUPPLEMENTAL INFORMATION

Supplemental information includes Supplemental Experimental Procedures, six figures, two tables, and one movie and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.008.

AUTHOR CONTRIBUTIONS

J.W. and J.L. expressed, purified, and grew crystals of the Cas1-Cas2-DNA complex. J.W., H.Z., and M.Y. collected X-ray diffraction data. J.L., J.W., G.S., and M.W. performed the biochemical assays. Y.W. solved the Cas1-Cas2-DNA complex structures, wrote the manuscript, and supervised all of the research.

ACKNOWLEDGMENTS

REFERENCES

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 513, 569–573.

Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, Ü. (2014). Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. Nucleic Acids Res. 42, 7884–7893.

Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., et al. (2011). A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol. Microbiol. 79, 484–502.

Barrangou, R., and Marraffini, L.A. (2014). CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. Mol. Cell 54, 234–244.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–1712.

Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., et al. (2008). A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. J. Biol. Chem. 283, 20361–20371.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008).

Cell

Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960–964.

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. 3, 945.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J. Bacteriol. 190, 1390–1400.

Díez-Villaseñor, C., Guzmán, N.M., Almendros, C., García-Martínez, J., and Mojica, F.J. (2013). CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli. RNA Biol. 10, 792–802.

Fineran, P.C., and Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. Virology 434, 202–209.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67–71.

Goren, M.G., Yosef, I., Auster, O., and Qimron, U. (2012). Experimental definition of a clustered regularly interspaced short palindromic duplicon in Escherichia coli. J. Mol. Biol. 423, 14–16.

Gunderson, F.F., Mallama, C.A., Fairbairn, S.G., and Cianciotto, N.P. (2015). Nuclease activity of Legionella pneumophila Cas2 promotes intracellular infection of amoebal host cells. Infect. Immun. 83, 1008–1018.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 139, 945–956.

Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., and Marraffini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature 519, 199–202.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J. Bacteriol. 190, 1401–1412.

Ka, D., Kim, D., Baek, G., and Bae, E. (2014). Structural and functional characterization of Streptococcus pyogenes Cas2 protein under different pH conditions. Biochem. Biophys. Res. Commun. 451, 152–157.

Kim, T.Y., Shin, M., Huynh Thi Yen, L., and Kim, J.S. (2013). Crystal structure of Cas1 from Archaeoglobus fulgidus and characterization of its nucleolytic activity. Biochem. Biophys. Res. Commun. 441, 720–725.

Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., and Sorek, R. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature 520, 505–510.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., et al. (2011). Evo-

lution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. 9, 467–477.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322, 1843–1845.

Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155, 733–740.

Nam, K.H., Ding, F., Haitjema, C., Huang, Q., DeLisa, M.P., and Ke, A. (2012). Double-stranded endonuclease activity in Bacillus halodurans clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. J. Biol. Chem. 287, 35943–35952.

Nuñez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., and Doudna, J.A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. Nat. Struct. Mol. Biol. 21, 528–534.

Nuñez, J.K., Lee, A.S., Engelman, A., and Doudna, J.A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature 519, 193–198.

Paez-Espino, D., Morovic, W., Sun, C.L., Thomas, B.C., Ueda, K., Stahl, B., Barrangou, R., and Banfield, J.F. (2013). Strong bias in the bacterial CRISPR elements that confer immunity to phage. Nat. Commun. 4, 1430.

Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L., and Severinov, K. (2010). Transcription, processing and function of CRISPR cassettes in Escherichia coli. Mol. Microbiol. 77, 1367–1379.

Rollie, C., Schneider, S., Brinkmann, A.S., Bolt, E.L., and White, M.F. (2015). Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. eLife 4, e08716.

Samai, P., Smith, P., and Shuman, S. (2010). Structure of a CRISPR-associated protein Cas2 from Desulfovibrio vulgaris. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 66, 1552–1556.

Swarts, D.C., Mosterd, C., van Passel, M.W., and Brouns, S.J. (2012). CRISPR interference directs strand specific spacer acquisition. PLoS ONE 7, e35888.

van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat. Rev. Microbiol. 12, 479–492.

Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S.M., Ma, W., and Doudna, J.A. (2009). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure 17, 904–912.

Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res. 40, 5569–5576.

Yosef, I., Shitrit, D., Goren, M.G., Burstein, D., Pupko, T., and Qimron, U. (2013). DNA motifs determining the efficiency of adaptation into the Escherichia coli CRISPR array. Proc. Natl. Acad. Sci. USA 110, 14396–14401.

Cell

Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System

Graphical Abstract
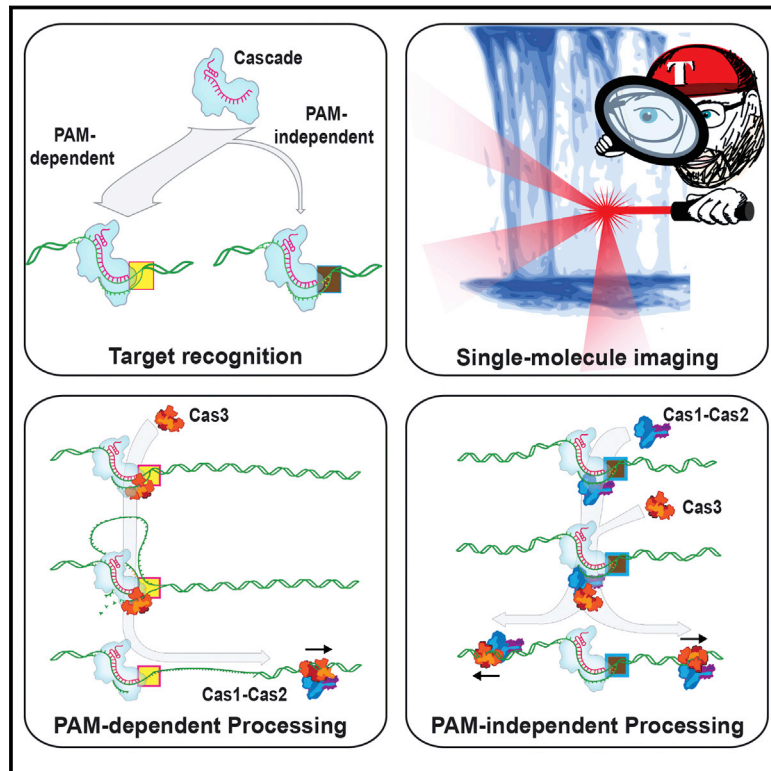
Authors
Sy Redding, Samuel H. Sternberg, Myles Marshall, ..., Blake Wiedenheft, Jennifer A. Doudna, Eric C. Greene

Correspondence
doudna@berkeley.edu (J.A.D.),
ecg2108@cumc.columbia.edu (E.C.G.)

In Brief
Single-molecule analysis of the bacterial Cascade complex reveals two distinct pathways leading to differential regulation of Cas3 DNA translocase and nuclease activities.

Highlights

- Cascade can locate targets through PAM-dependent and PAM-independent pathways

- PAM-dependent recognition enables direct recruitment of the Cas3 translocase/nuclease

- PAM-independent recognition requires Cas1-Cas2 for Cas3 recruitment

- Cas1-Cas2 serve as *trans*-acting factors that regulate Cas3 activities

CrossMark

CellPress

Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System

Sy Redding,[1,9] Samuel H. Sternberg,[2] Myles Marshall,[8] Bryan Gibb,[8,10] Prashant Bhat,[3,11] Chantal K. Guegler,[2,12] Blake Wiedenheft,[4] Jennifer A. Doudna,[2,3,5,6,7,*] and Eric C. Greene[8,*]

[1]Department of Chemistry, Columbia University, New York, NY 10027, USA
[2]Department of Chemistry, University of California, Berkeley, CA 94720, USA
[3]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA
[4]Department of Microbiology and Immunology, Montana State University, Bozeman, MT 59717, USA
[5]Innovative Genomics Initiative, University of California, Berkeley, CA 94720, USA
[6]Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA
[7]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[8]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA
[9]Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94143, USA
[10]Present address: Department of Life Sciences, New York Institute of Technology, Old Westbury, NY 11568, USA
[11]Present address: David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA
[12]Present address: Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA
*Correspondence: doudna@berkeley.edu (J.A.D.), ecg2108@cumc.columbia.edu (E.C.G.)
http://dx.doi.org/10.1016/j.cell.2015.10.003

SUMMARY

CRISPR-Cas adaptive immune systems protect bacteria and archaea against foreign genetic elements. In *Escherichia coli,* Cascade (CRISPR-associated complex for antiviral defense) is an RNA-guided surveillance complex that binds foreign DNA and recruits Cas3, a *trans*-acting nuclease helicase for target degradation. Here, we use single-molecule imaging to visualize Cascade and Cas3 binding to foreign DNA targets. Our analysis reveals two distinct pathways dictated by the presence or absence of a protospacer-adjacent motif (PAM). Binding to a protospacer flanked by a PAM recruits a nuclease-active Cas3 for degradation of short single-stranded regions of target DNA, whereas PAM mutations elicit an alternative pathway that recruits a nuclease-inactive Cas3 through a mechanism that is dependent on the Cas1 and Cas2 proteins. These findings explain how target recognition by Cascade can elicit distinct outcomes and support a model for acquisition of new spacer sequences through a mechanism involving processive, ATP-dependent Cas3 translocation along foreign DNA.

INTRODUCTION

Many prokaryotes harbor an RNA-guided adaptive immune system comprised of a genetic locus called CRISPR (clustered regularly interspaced short palindromic repeats) and the CRISPR-associated (Cas) genes (Barrangou and Marraffini, 2014; van der Oost et al., 2014; Westra et al., 2012a). The CRISPR locus was first identified in *Escherichia coli* as an un-

usual series of 29-bp repeats separated by 32-bp spacer sequences (Ishino et al., 1987). It was later recognized that these spacers were derived from foreign genetic elements, suggesting the CRISPR locus might serve as an RNA-guided immune system (Bolotin et al., 2005; Makarova et al., 2006; Mojica et al., 2005). It is now known that CRISPR-Cas immunity is conferred through integration of short DNA fragments into the CRISPR locus, and these spacer sequences record the history of past infections (Barrangou and Marraffini, 2014; van der Oost et al., 2014; Westra et al., 2012a). The CRISPR locus is transcribed, and the resultant transcript is processed into shorter CRISPR-RNAs (crRNAs), each containing a sequence complementary to a previously encountered foreign DNA element.

CRISPR-Cas systems are classified as types I, II or III, which can be distinguished based on the presence of the signature *Cas3, Cas9,* or *Cas10* genes, respectively (Barrangou and Marraffini, 2014; Westra et al., 2012a). Type I are the most common, and much of our understanding of type I CRISPR-Cas systems comes from studies of *E. coli* Cascade (CRISPR-associated complex for antiviral defense), which is comprised of the five proteins Cse1, Cse2, Cas7, Cas5e, and Cas6e. These proteins assemble on a 61-nt crRNA, yielding a 405- kDa complex. The crRNA contains the 32-nt spacer sequence, which directs Cascade to sequences (protospacers) in foreign DNA, leading to formation of an R-loop intermediate. Cascade then recruits Cas3, which has an N-terminal histidine-aspartate (HD) nuclease domain and C-terminal superfamily 2 (SF2) helicase domain, to degrade the DNA (Mulepati and Bailey, 2013; Sinkunas et al., 2013).

Cascade must discriminate between spacer sequences found in the bacterial chromosome and those found in foreign DNA. This discrimination is thought to be accomplished through recognition of a trinucleotide sequence motif called the protospacer-adjacent motif (PAM; 5′-A[A/T]G-3′ for *E. coli* Cascade), which is adjacent to the protospacer in foreign DNA, but absent in the CRISPR locus. Strict sequence
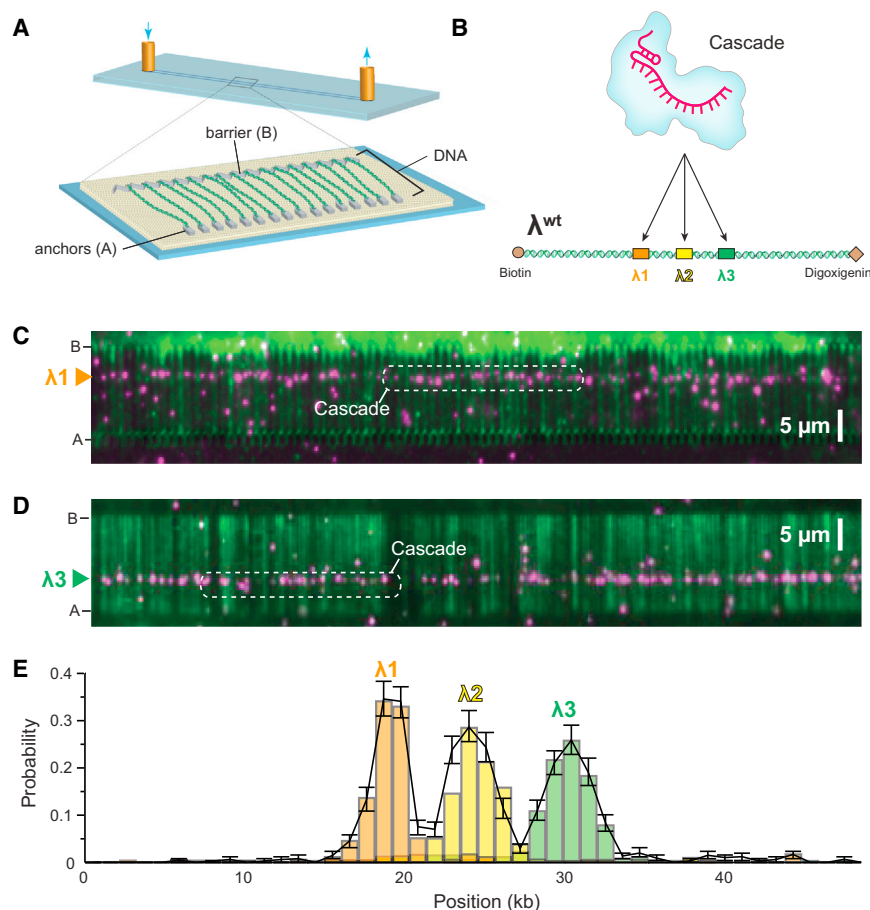
Figure 1. Programmed Target Binding by _E. coli_ Cascade

(A) Overview of DNA curtains.

(B) Schematic of _E. coli_ Cascade programmed with a crRNA targeting one of three different binding sites (designated λ1, λ2, and λ3) on λ^{WT}.

(C) Wide-field TIRF microscopy image showing QD-tagged Cascade (magenta) bound to DNA (green) at λ1.

(D) Wide-field image showing Cascade bound at λ3.

(E) Binding distribution for Cascade targeted to each of the three protospacers; error bars here and all subsequent binding distributions represent 95% confidence intervals obtained through bootstrap analysis.

Cascade can still bind tightly to the DNA, ensuring that it can initiate the sequence of molecular events that precede primed spacer acquisition. Through this pathway, Cas3 recruitment becomes strictly dependent on Cas1-Cas2, and Cas1-Cas2 also attenuate Cas3 nuclease activity and enable Cas3 to rapidly translocate in either direction along the foreign DNA. These results establish Cas1-Cas2 as a _trans_-acting factor necessary for the recruitment and regulation of Cas3 at escape targets. Based on our findings, we propose a mechanistic framework describing how Cascade, Cas1, Cas2, and Cas3 work together to process and disable foreign genetic elements.

requirements present a potential weakness because mutations in either the PAM or protospacer can allow foreign DNA to escape CRISPR-Cas immunity (Semenova et al., 2011). However, bacteria can rapidly restore immunity using a positive-feedback loop to update the CRISPR locus (Datsenko et al., 2012; Fineran et al., 2014). The mechanism of primed spacer acquisition (priming) remains perhaps one of the most poorly understood aspects of CRISPR-Cas immunity (Datsenko et al., 2012; Fineran et al., 2014; Heler et al., 2014). Priming requires Cascade with a crRNA bearing at least partial complementarity to the escape target, suggesting Cascade must be able to locate targets even when they bear mutations sufficient to escape immunity (Datsenko et al., 2012). Priming also requires Cas3 (Datsenko et al., 2012) and the Cas1-Cas2 complex (Nuñez et al., 2014), which integrate new sequences into the CRISPR locus (Nuñez et al., 2015). It is not known how these complexes elicit the priming response to foreign elements bearing escape mutations.

Here, we use single-molecule imaging to visualize individual Cascade complexes as they search for protospacers within the bacteriophage λ genome. Our work reveals PAM-dependent and PAM-independent search pathways. The PAM-dependent pathway is highly efficient and allows Cascade to recruit Cas3 for strand-specific degradation of the target genome. The PAM-independent pathway is less efficient, but

RESULTS

DNA Curtain Assay for Target Binding by Cascade

We sought to establish a DNA curtain assay using total internal reflection fluorescence (TIRF) microscopy for visualizing the behavior of Cascade on individual molecules of wild-type phage λ DNA (λ^{WT}) (Figure 1A; Supplemental Experimental Procedures) (Greene et al., 2010). In brief, the surface of a microfluidic sample chamber was coated with a lipid bilayer, and DNA molecules were anchored to the bilayer through a biotin-streptavidin interaction. The DNA was then pushed to the leading edges of nanofabricated barriers to lipid diffusion, and the downstream ends were anchored to pedestals through antibody-hapten linkage (Gorman et al., 2010, 2012). Cascade was prepared with one of three crRNAs targeted to different regions of λ^{WT} and then labeled with antiFLAG-quantum dots (QDs) attached to the 3xFLAG-tagged Cas6e subunit (Figure 1B). When visualized on DNA curtains, Cascade bound to target sites corresponding to DNA sequences complementary to the three different crRNAs (Figures 1C–1E). Cascade remained bound for at least 57 min; this lifetime represents a lower limit for the Cascade-protospacer interaction because these measurements are limited by the
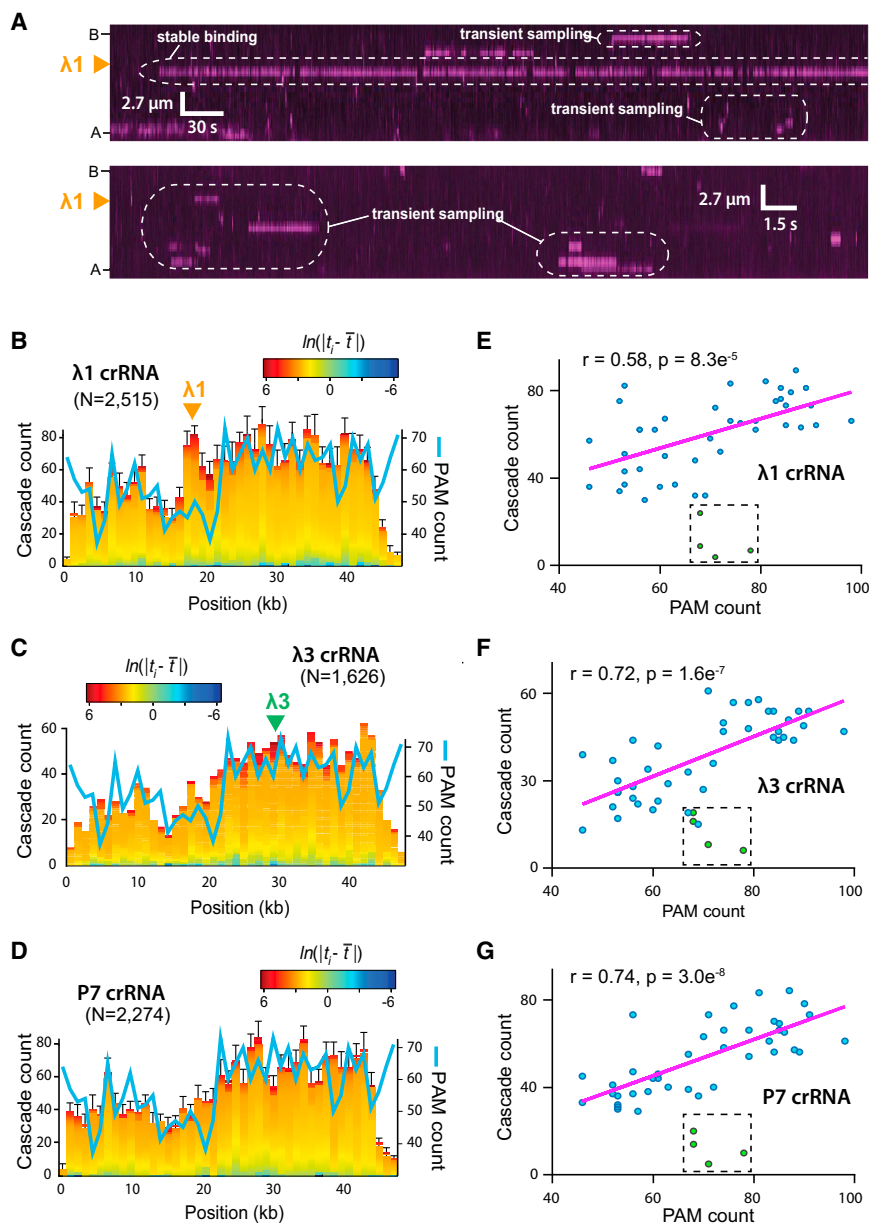
Figure 2. Cascade Searches for PAMs while Interrogating Foreign DNA

(A) Kymographs highlighting examples of Cascade binding events over two different time regimes (see scale bars). Examples of transient sampling and stable recognition are highlighted.

(B–D) Distribution of PAMs (blue line) and transient binding events for Cascade programmed with (B) the λ1-crRNA, (C) the λ3-crRNA, or (D) a P7-crRNA. Count refers to number of occurrences within 1 kbp of DNA. The locations of the λ1 and λ3 target sites are indicated, and the heat map color-coding reflects the binding dwell time (t_i) relative to the mean dwell time ($\bar{t}$).

(E–G) Correlation of PAMs with the transient binding events for Cascade programmed with (E) the λ1-crRNA, (F) the λ3-crRNA, or (G) P7-crRNA, as indicated. Outlying data points (colored green and boxed) reflect underrepresented binding events at PAM sites near the ends of the DNA; detection of binding at these sites is hindered by the chromium barriers.

See also Figure S1.

~250 bp, although we do not exclude that possibility that Cascade may diffuse shorter distances along the DNA (Gorman et al., 2010, 2012). The remaining fraction of Cascade molecules (<25%) underwent optically detectable 1D diffusion; we did not pursue a detailed analysis of this behavior because it coincided with a loss of binding specificity and appeared to arise from Cascade aggregates (data not shown).

Analysis of the 3D events revealed long-lived binding to the protospacers, as well as transient binding events all along the λ^WT DNA (Figure 2A). The λ^WT genome contains a total of 3,151 PAM sites (5′-A[A/T]G-3′), corresponding to ~1 PAM per 15.4 bp, which are asymmetrically distributed across the phage genome. Cascade did not randomly sample the DNA, instead the transient binding events were correlated with the PAM distribution (Figures 2E–2G), as we have reported for *Streptococcus pyogenes* Cas9 (Sternberg et al., 2014). Control reactions using Cascade programmed with P7-crRNA revealed a similar pattern of transient binding (Figures 2B–2G), and we could detect no binding activity for Cascade lacking Cse1 (data not shown), which is the subunit responsible for PAM recognition (Sashital et al., 2012). Cascade programmed with either λ1-crRNA or λ3-crRNA displayed many reversible binding events at their targets, which are revealed by the ~50% increased prevalence of longer-lived intermediates at both of these target sites relative to non-target sites (Figures 2B, 2C, and S1A), and also by the peak in binding at λ1 for the λ1-crRNA, which is observable due to the overall lower density of PAM sites in this region of DNA (Figures 2B and 2C). This

stability of the 3xFLAG-antiFLAG interaction (Sternberg et al., 2014). Stable binding was not observed for Cascade bearing a control crRNA (P7-crRNA) that was not complementary to λ^WT.

PAM-Dependent Target Recognition

Next, we sought to determine how Cascade locates protospacers by visualizing reactions in real-time. Most Cascade (> 75%) appeared immediately at the protospacer without exhibiting any evidence of microscopically detectable motion along the DNA (Figure 2A). This finding leads us to conclude that Cascade located the protospacer through a pathway that was dominated by 3D diffusion at the microscopic scale. Based on our optical resolution limits, these experiments provide an upper limit on any potential 1D diffusion by Cascade of no more than
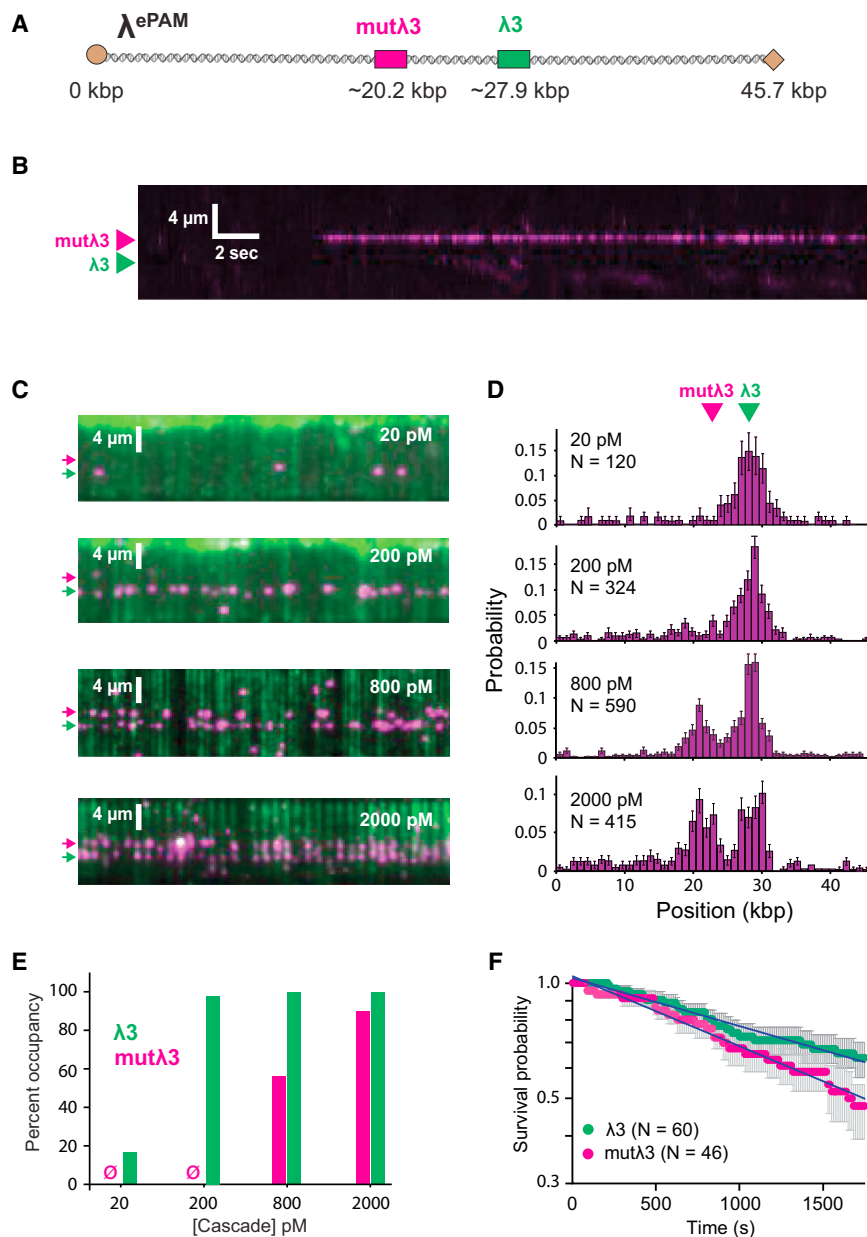
Figure 3. Recognition of Escape PAM Mutants

(A) Schematic of λ^ePAM bearing two identical protospacers, one with a cognate PAM (λ3) and the other with an escape PAM (mutλ3).

(B) Kymograph highlighting example of Cascade binding to the mutλ3 through 3D diffusion.

(C) Wide-field images showing binding to each of the two targets at different Cascade concentrations following a 10-min incubation. Arrowheads indicate the locations of the λ3 (green) and mutλ3 (magenta) targets.

(D) Binding distributions showing relative occupancy at each Cascade concentration.

(E) Quantification of percent occupancy; Ø indicates no detectable binding.

(F) Survival probability plots for Cascade bound to the two targets; error bars here and all subsequent survival probability plots represent 70% confidence intervals obtained through bootstrap analysis.

See also Table S1.

association, provide further evidence that the initial observed interactions are based on a sequence-dependent association with PAM sites rather than on nonspecific interactions with the DNA phosphate backbone.

PAM-Independent Target Recognition

Next, we sought to determine whether and how Cascade locates targets that lack a canonical PAM. For this, we generated a new phage construct (λ^ePAM) bearing two duplicate targets (Figure 3A). One of the protospacers (λ3) was adjacent to a cognate PAM [5′-ATG-3′], whereas second protospacer (mutλ3) was adjacent to a mutated PAM [5-ATT-3′]. This escape PAM (ePAM) was chosen because it enables an invading DNA to escape the CRISPR-Cas machinery, but still elicits a rapid priming response (Datsenko et al., 2012; Fineran et al., 2014).

category of long-lived, but reversible, binding events at the protospacers likely represents abortive engagement, suggesting Cascade must often make multiple attempts before stably engaging the protospacer, similar to what we have observed for Cas9 (Sternberg et al., 2014).

The transient binding events exhibited double-exponential decays similar to *S. pyogenes* Cas9 (Sternberg et al., 2014), with lifetimes of ~3 and ~25 s (Figures S1A–S1D), indicating that at least two intermediates exist on the pathway toward target recognition. The lifetimes of these intermediates were not appreciably affected by either salt concentration or temperature (Figure S1D), similar to findings for Cas9 (Sternberg et al., 2014). These characteristics, more commonly attributed to site-specific

Surprisingly, Cascade could still bind both protospacers, and binding of mutλ3 still occurred through 3D diffusion (Figure 3B), but recognition of mutλ3 was much less efficient than recognition of λ3 (Figures 3C–3E). This difference was evidenced by the ~10-fold higher Cascade concentration necessary to achieve similar levels of occupancy at both protospacers. Despite the large difference in initial recognition, the lifetimes of Cascade at λ3 and mutλ3 were comparable (57 and 40 min, respectively; Figure 3F). We conclude that PAMs increase the efficiency of target recognition, but that Cascade is still capable of protospacer recognition and high-affinity binding in the absence of a cognate PAM, and this conclusion is consistent with previous studies (Szczelkun et al., 2014).

A

B

4 µm   RPA foci   RPA-eGFP (+Cas3)

λ3 ▶

A

B

B

4 µm

λ3 ▶

A   Cascade (–Cas3)

B

4 µm

λ3 ▶

A   RPA foci (–Cas3)

C

RPA signal (y-axis: 0, 10, 20, 30, 40, 50)

ATP	–	+	+	+	+
Cas3	+	+	–	+	–
Cas3-D75A	–	–	+	–	–
Co²⁺	+	+	+	–	+

D

λ3

RPA-eGFP foci (N = 564)

Count (y-axis: 0, 10, 20, 30, 40, 50)

Position (kbp) (x-axis: 0, 10, 20, 30, 40)

Intensity (a.u.): 210, 150, 90, 30

E

~3 x RPA-GFPs

Probability (y-axis: 0, 20, 40, 60)

RPA-eGFP foci (N = 546)

Intensity (a.u.) (x-axis: 0, 50, 100, 150, 200, 250)

F

3 RPA-eGFP
2 RPA-eGFP
1 RPA-eGFP

Signal (a.u.) (y-axis: 0, 10, 20, 30)

Time (s) (x-axis: 20, 40, 60)

Figure 4. Cas3 Generates an ssDNA Gap at the λ3 Protospacer

(A) Image showing RPA-eGFP foci at λ3 for reactions with unlabeled Cascade and unlabeled Cas3.

(B) Control images showing that RPA-eGFP foci are not present when Cas3 is omitted from the reactions; the upper and lower panels show the same field of view.

(C) Requirements for RPA-eGFP foci formation at λ3.

(D) Distribution of RPA-eGFP foci in reactions containing both Cascade and Cas3; Count refers to the number of occurrences within 1 kbp of DNA.

(E) Signal intensities for RPA-eGFP foci. The intensity of a focus comprised of three molecules of RPA-eGFP is indicated, and each successive bin corresponds to ~1 additional molecule of RPA-eGFP. The heat map color-coding in (D) and (E) are the same.

(F) Representative stepwise photobleaching curve used to estimate the number of RPA-eGFP molecules in each focus.

See also Figure S2.

through an ATP-, Mg²⁺-, and Co²⁺-dependent mechanism (Mulepati and Bailey, 2011, 2013; Sinkunas et al., 2013). Cas3 degrades both target DNA strands in bulk biochemical assays (Mulepati and Bailey, 2013; Sinkunas et al., 2013). However, these measurements use relatively high concentrations of Cas3 (50 nM–1 µM) (Hochstrasser et al., 2014; Mulepati and Bailey, 2011, 2013; Sinkunas et al., 2011, 2013), suggesting that DNA degradation may be due to the action of multiple Cas3 molecules, only the first of which is directly recruited by Cascade. Given these considerations, it is plausible that the initial Cascade-recruited molecule of Cas3 only introduces a small nick or ssDNA gap in the target DNA (Mulepati and Bailey, 2013).

We reasoned that if Cas3 was initially generating ssDNA after loading at Cascade, then this might be revealed in reactions with low concentrations of Cas3 (4 nM), followed by the addition of eGFP-tagged replication protein A (RPA), which binds ssDNA. When RPA-eGFP was added after Cascade and Cas3, bright eGFP foci were detected at the λ3 protospacer (Figure 4A). Formation of RPA-eGFP foci was dependent on Cascade, Cas3, ATP, and Co²⁺ and the conditions under which we detected RPA-eGFP foci paralleled the conditions necessary for plasmid degradation in bulk biochemical assays (Figures 4B and S2C). Furthermore, RPA-eGFP foci were not observed for a Cas3 nuclease mutant (D75A) (Figures 4C and S2A–S2C). Notably, the DNA in the single-molecule assays

Cas3 Recruitment Leads to Disruption of the Target DNA Duplex

Next, we sought to visualize Cascade-dependent recruitment of Cas3. Cas3 interacts with Cse1, and the displaced single-stranded DNA (ssDNA) strand that is generated by R-loop formation (Hochstrasser et al., 2014; Mulepati and Bailey, 2013; Sinkunas et al., 2013). Upon recruitment, Cas3 first nicks the DNA and is thought to then translocate in the 3′ → 5′ direction along the non-target strand, while unwinding and degrading duplex DNA

Figure 5. Cascade-Mediated Recruitment of Cas3

(A) Image showing that QD-tagged Cas3 is recruited to unlabeled Cascade at λ3.

(B) Binding of Cas3 to λ3. The distribution is segregated into the translocation (orange) and stationary (green) Cas3 populations.

(C) Survival probabilities of the stationary Cas3 population.

(D) Kymograph illustrating the translocation of Cas3 away from λ3 in a reaction with unlabeled Cascade. The delay period prior to the initiation of Cas3 translocation is indicated.

(E) Two-color experiment showing that Cas3 (green) translocates away from Cascade (magenta).

(F) Survival probability (delay time) of the translocating population of Cas3 prior to moving away from λ3.

(G) Cas3 velocity distribution.

(H) Cas3 processivity distribution.

(I) Kymograph showing an example of Cas3 repeatedly looping the DNA.

(J) Intensity profile showing the increase in Cas3 fluorescence signal coinciding with DNA loop formation.

See also Figure S3 and Movie S1.

Cas3 Recruitment to Target-Bound Cascade

Next, we sought to visualize the behavior of fluorescently tagged Cas3 (Figure S3A; Supplemental Experimental Procedures). We were unable to detect stable binding of Cas3 to Cascade when ATP or Mg^{2+} were omitted or when ATP was replaced with ADP or AMP-PNP (data not shown). However, Cas3 bound stably to Cascade when ATP and Mg^{2+} were included in the reactions (Figures 5A and 5B). Cas3 located Cascade through 3D diffusion during initial recruitment (see Figure 5D). Once bound, ~55% of the Cas3 molecules remained stationary within optical resolution limits (Figure 5B). These seemingly stationary molecules exhibited two distinct lifetimes: one population with a lifetime (τ_1) of ~6 s and a second population (τ_2) with a lifetime of ≥ 1 min (Figures 5C, S3B, and S3C). These findings suggest that Cas3 transiently samples target-bound Cascade before transitioning into a more stably bound state and that entry into this longer-lived state requires ATP hydrolysis. Interestingly, once a longer-lived Cas3 binding event was observed at a given molecule of Cascade, then that particular Cascade complex appeared incapable of recruiting any additional Cas3 at the protein concentrations used in these assays.

Cas3 Is a Highly Processive Molecular Motor

Many of the Cas3 molecules (~45%) translocated along the DNA (Figures 5B, 5D, and 5E). In these instances, Cas3 was recruited

was not liberated from the flow cell surface and there was no evidence for long tracts of RPA-eGFP, indicating that Cas3 only generated a small ssDNA gap. To estimate the size of the ssDNA gaps, we measured the intensity of the RPA-eGFP foci (Figures 4D and 4E) and then used photobleaching steps to roughly estimate the number of RPA-eGFP molecules present (Figure 4F; Supplemental Experimental Procedures). We estimate that the average focus contained ~8–10 molecules of RPA-eGFP, corresponding to ~240–300 nt of ssDNA. These results suggest that the first Cas3 molecule recruited by Cascade makes a short ssDNA gap adjacent to the protospacer.

to Cascade at the λ3 protospacer and then moved rapidly away from the protospacer in a direction consistent with $3' \rightarrow 5'$ translocation on the non-target strand, as expected from bulk biochemical experiments (Mulepati and Bailey, 2013). There was no evidence that Cas3 translocation could initiate from any other location on the DNA other than the λ3 protospacer, and Cas3 translocation was entirely dependent on the presence of Cascade. Remarkably, Cascade remained tightly bound to the protospacer even after Cas3 had begun translocating along the DNA (Figure 5E). Moreover, once Cas3 had translocated away from Cascade, then no additional molecules of Cas3 could bind to or translocate away from that particular Cascade complex.

Cas3 exhibited a short delay prior to moving away from Cascade (Figure 5D); analysis of these delay times revealed two lifetimes that were similar to the τ_1 and τ_2 lifetimes for the stationary Cas3 population, suggesting that the observed intermediates reflected the same underlying molecular processes (Figures 5C, 5F, S3B, and S3C). Cas3 traveled at a mean velocity of ~316 bp/s for 12,000 bp before stalling or dissociating from the DNA (Figures 5G and 5H), and >99% of molecules exhibited unidirectional movement (Figures 5D and 5E; Movie S1; see below). Three key observations suggested that Cas3 was not extensively degrading the DNA during translocation. First, there was no evidence that the translocating population of Cas3 caused double-strand breaks. Second, we saw no evidence for long ssDNA tracts when reactions were chased with RPA-eGFP. Finally, if Cas3 had generated tracts of ssDNA long enough to be optically detected, then Cascade would also appear to move in the same direction because of the change in persistence length that accompanies the conversion of dsDNA to ssDNA, but Cascade always remained stationary at the protospacer. We conclude that Cas3 is a highly processive molecular motor that first generates a small ssDNA gap and then translocates in $3' \rightarrow 5'$ direction along the non-target DNA strand away from Cascade.

Evidence for Looped DNA Intermediates

Surprisingly, in addition to our observation that Cas3 recruitment and translocation did not coincide with the ejection of Cascade from the DNA, inspection of the Cas3 translocation trajectories revealed evidence that the contacts between Cas3 and Cascade were not immediately broken. In many instances (14%), Cas3 began to translocate along the DNA, but then returned almost instantaneously to the original binding site (Figure 5I). This behavior coincided with an increase in Cas3 fluorescence, suggesting that the molecules were pulled closer to the surface of the flow cell because of increased tension on the DNA. These observations are most consistent with looped DNA intermediates, where Cas3 maintains contact with Cascade, while simultaneously translocating for a short distance along the flanking duplex DNA (Figure S3D). We conclude that Cas3 can initially remain bound to Cascade as it begins translocating along the DNA and that a subset of these molecules generates optically detectable DNA loops.

PAM Is Essential for Cascade-Mediated Recruitment of Cas3

Next, we sought to determine whether Cascade could recruit Cas3 to mutλ3, and if so, whether the properties of Cas3 differ in the absence of a cognate PAM. Interestingly, Cas3 did not co-localize with Cascade at mutλ3 (Figure S4A), and we were unable to detect even transient binding of Cas3 to Cascade at the mutλ3 protospacer. We were also unable to detect RPA-eGFP foci at mutλ3 (Figures S4B–S4E), and Cas3 did not cleave plasmid substrates bearing the mutλ3 protospacer (see below). We conclude that Cascade cannot recruit Cas3 to DNA in the absence of a cognate PAM, in agreement with previous bulk biochemical experiments (Hochstrasser et al., 2014; Mulepati and Bailey, 2013).

PAM-Independent Recruitment of Cas3 by Cas1-Cas2

Cas3 is required for primed sequence acquisition (Datsenko et al., 2012), suggesting that alternative pathways must exist to recruit Cas3 to escape targets. Cas1 and Cas2 are universally conserved across CRISPR types and are also necessary for primed sequence acquisition, suggesting the possibility that these proteins may work in concert with Cascade to promote the recruitment of Cas3 to escape targets. Therefore, we next asked whether the Cas1-Cas2 complex might affect target recognition, target processing, or both, in reactions with Cas3. Attempts to generate fluorescently tagged Cas1 or Cas2 yielded inactive proteins, and therefore these experiments utilized wild-type (unlabeled) Cas1-Cas2.

Remarkably, the addition of Cas1-Cas2 enabled the recruitment of Cas3 to mutλ3 and also ~3-fold enhanced recruitment of Cas3 to λ3 (Figures 6A and 6B; Movie S1 and S2). The velocity and processivity of Cas3 were not altered by Cas1-Cas2 (Figures S5A and S5B). However, Cas3 recruited to the escape target behaved markedly different from Cas3 that was recruited to cognate protospacer. Most strikingly, Cas3 targeted to mutλ3 could rapidly translocate in either direction away from Cascade (Figure 6C; Movie S3). Moreover, Cas3 exhibited only a ~6 s delay prior to moving away from mutλ3, but there was no evidence for the second longer-lived intermediate (τ_2) that was always observed at λ3 (Figures S3B, S3C, and S5C). There was also no evidence for ssDNA gaps at mutλ3 in the presence of Cas1-Cas2 (Figure S5D), and bulk biochemical assays with Cascade, Cas1-Cas2, and Cas3 revealed no nicking or cleavage of plasmids with the mutλ3 protospacer (Figures S6A and S6B), even though Cascade was capable of binding the mutλ3 protospacer in bulk assays (Figure S6C). Finally, there was no evidence for Cas3-mediated DNA looping at mutλ3 in reactions with Cas1-Cas2 (Figure 6C). Together, these results show that Cas1-Cas2 are necessary to recruit Cas3 to mutλ3 and attenuate the nuclease activity of Cas3 at these escape targets, enabling Cas3 to translocate away from Cascade in either direction along the foreign DNA.

Cas1-Cas2 also appeared to affect the behavior of Cas3 at the λ3 protospacer. Specifically, Cas1-Cas2 partially attenuated Cas3 nuclease activity in bulk biochemical assays (Figures S6A and S6B), and the presence of Cas1-Cas2 also enabled iterative Cas3 binding and translocation events from the same Cascade complex bound to the λ3 protospacer (Movie S4). This observation was in stark contrast to reactions done in the absence of Cas1-Cas2, where we never detected evidence of multiple Cas3 recruitment events to the same Cascade complex. These findings suggest that Cas1-Cas2 not only enhances the recruitment of Cas3 to Cascade bound at λ3 but may also enable iterative Cas3 loading events.
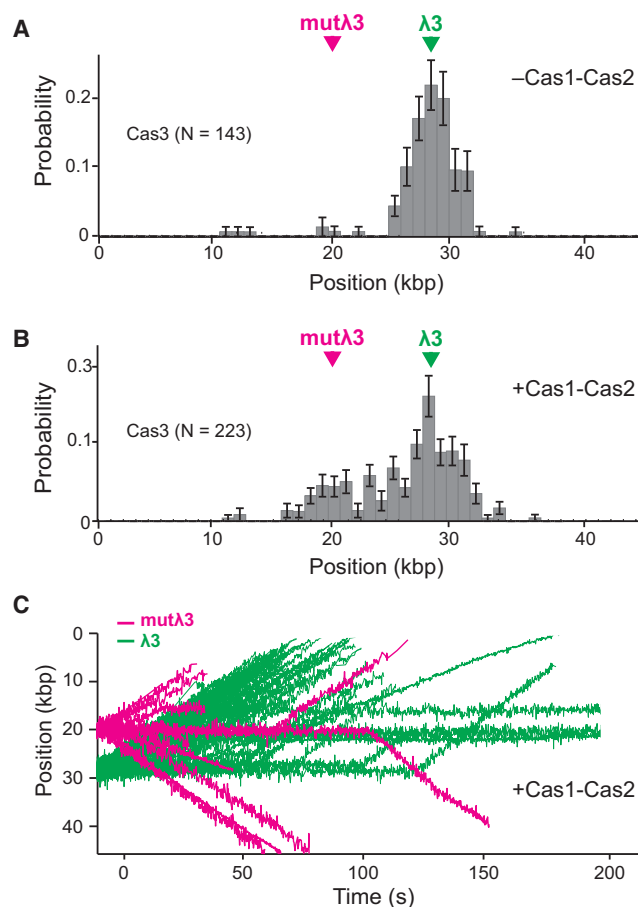
Figure 6. Cas1-Cas2-Mediated Recruitment of Cas3 to Escape Targets

(A) Binding distribution of Cas3 on λ^{ePAM} in the absence of Cas1-Cas2.

(B) Cas3 binding distribution histogram on λ^{ePAM} in the presence of Cas1-Cas2.

(C) Overlaid trajectories showing examples of Cas3 translocation events originating from either the λ3 protospacer (green) or the mutλ3 protospacer (magenta). Of the trajectories originating from mutλ3, 59% of the Cas3 molecules move toward the downstream anchor points, and the remaining 41% travel in the opposite direction.

See also Figures S4 and S5 and Movies S2, S3, and S4.

DISCUSSION

CRISPR-Cas immunity involves complex interplay among multiple macromolecular components, with the potential for overlapping or convergent pathways. Our work reveals two distinct pathways for target recognition and processing and shows that the choice of pathway is dictated by the presence or absence of a PAM sequence adjacent to the targeted protospacer (Figure 7).

A Conserved Mechanism for PAM-Dependent Target Recognition

Our results support a model in which an initial search for PAM sequences is the predominant mode of DNA surveillance by *E. coli* Cascade (Figure 7A). Once a PAM is identified, Cascade interro-

gates the flanking DNA for sequence complementarity to the crRNA via directional unwinding of the DNA beginning at the PAM, and identification of a matching protospacer leads to stable capture and R-loop formation (Rutkauskas et al., 2015). This PAM-dependent search process is strikingly similar to that of *S. pyogenes* Cas9, the crRNA-guided surveillance complex in type II CRISPR-Cas systems, which also initiates the search by looking for PAMs (Sternberg et al., 2014). In addition, the type IF CRISPR-Cas system of *Pseudomonas aeruginosa* also searches for PAM sequences before probing the flanking DNA for sequence complementarity to the crRNA (Rollins et al., 2015). The type II CRISPR-Cas systems require only a single polypeptide for target recognition and cleavage, whereas type I CRISPR-Cas systems require large multimeric complexes for target recognition and a separate *trans*-acting protein (Cas3) for DNA cleavage. Cas9 and Cse1 share no amino acid sequence homology, and the Cas9 PAM (5′-NGG-3′) and the Cascade PAM (5′-A[A/T]G-3′) are located on opposite ends of the protospacer and on different DNA strands (Jinek et al., 2012; Sashital et al., 2012). Given these differences, there was no reason to assume that *S. pyogenes* Cas9 and *E. coli* Cascade would search for target sites using the same general mechanism. The similarities between Cascade and Cas9 suggest that an initial search for PAMs may be a broadly conserved mechanism for DNA surveillance among the type I and type II CRISPR-Cas systems.

Facilitated Diffusion versus Reduced Complexity

It is often assumed that site-specific DNA binding proteins accelerate target searches relative to 3D diffusion by facilitated diffusion, which reduces the dimensionality of the search process through 1D sliding, hopping, and/or intersegmental transfer (von Hippel and Berg, 1989). However, there is little evidence supporting this general assumption (Halford, 2009). The Cascade target search is remarkably similar to that of Cas9's, which also exhibits no evidence of 1D sliding (Sternberg et al., 2014). Instead, we find that Cascade and Cas9 both appear to optimize their target searches by reducing the complexity of the sequence space that is sampled while surveying DNA. They accomplish this task by first looking for a small portion of the overall binding site, the PAM, before probing the flanking DNA for sequences complementary to the crRNA, which provides an additional layer of discrimination enabling Cascade to sample and reject incorrect targets (Rutkauskas et al., 2015; Sternberg et al., 2014). The effectiveness of this strategy can be illustrated by considering that based on sequence composition alone Cascade can avoid ~90% of the λ genome just by utilizing the PAM as an initial recognition signal, while kinetically ignoring other sequences. The finding that much higher Cascade concentrations are necessary to achieve similar occupancy at protospacers with an escape PAM compared to those with a cognate PAM also reflects the effectiveness of reducing search complexity.

PAM-Dependent Target Processing

The PAM-dependent pathway requires only Cascade to recruit Cas3 to protospacers (Figure 7B). Cas3 first transiently samples Cascade before transitioning into a stably bound complex.
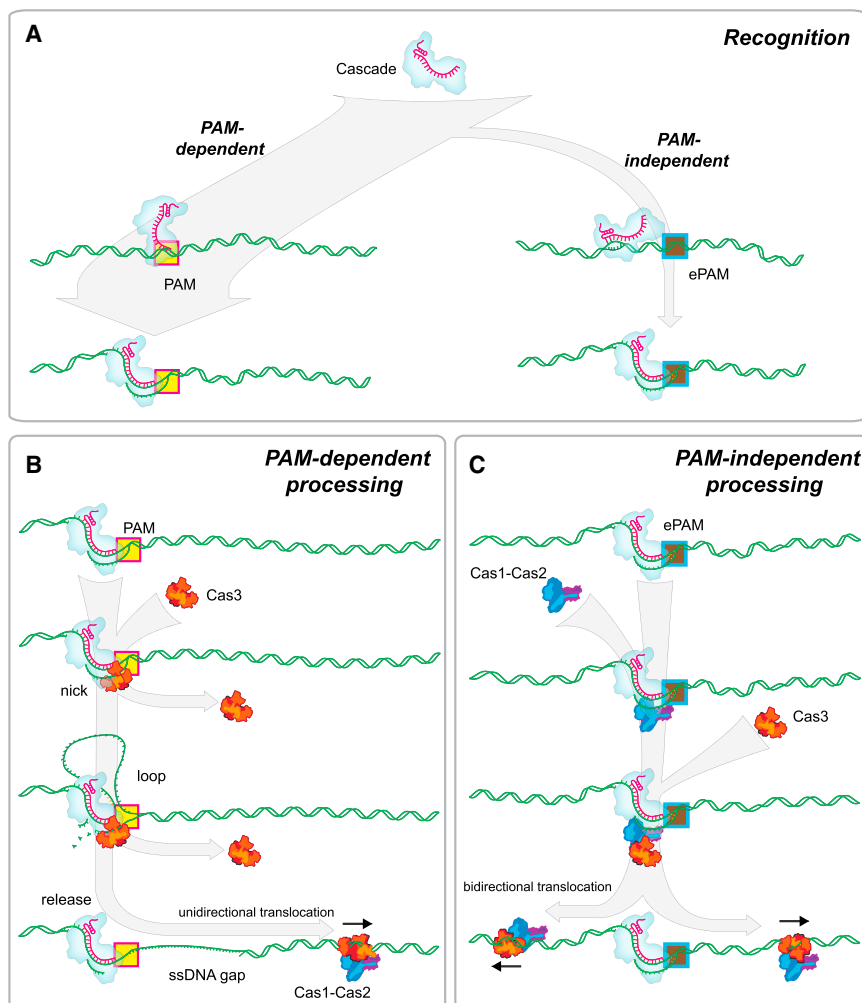
Figure 7. Model for Foreign DNA Recognition and Processing by Cascade, Cas1, Cas2, and Cas3

(A) The predominant mechanism for protospacer recognition is through the PAM-dependent pathway.

(B) PAM-dependent processing involves the recruitment of Cas3 to the protospacer by Cascade. Cas3 nicks the R-loop and generates an ssDNA gap; Cas3 can dissociate at either of these two steps. Cas3 then breaks free from Cascade and travels unidirectionally along the DNA.

(C) PAM-independent processing requires Cas1-Cas2 to recruit Cas3. Cas3 is loaded onto the DNA in one of two possible orientations through a mechanism that attenuates Cas3 nuclease activity. Cas3 then travels in either direction along the DNA as part of a spacer acquisition complex.

See also Figure S6.

suggest that the early stages of foreign DNA degradation involve the ATP-dependent recruitment of just one molecule of Cas3 through a mechanism that requires Cascade-specific contacts and an intact R-loop. This initial transient binding event exhibits a ~6-s lifetime (τ_1) before Cas3 transitions into a more stably bound intermediate. The first stably bound molecule of Cas3 then generates a short ssDNA gap, reflected in the delay time (τ_2) prior to moving away from Cascade, and after being released from Cascade, this Cas3 molecule can either dissociate into solution or continue traveling along the remaining duplex DNA. Any subsequent recruitment of Cas3 (or other nucleases) occurs through nonspecific interactions with the resultant ssDNA gap.

Cascade remains tightly bound to the DNA even after Cas3 generates an ssDNA gap and moves away from the protospacer. It is possible that continued presence of Cascade may distinguish these Cas3-generated gaps from other ssDNA gaps that can be produced during normal DNA metabolism, and Cascade may perhaps prevent host DNA repair proteins from filling in these gaps before the invading DNA is eventually destroyed.

PAM-Dependent Cas3 Motor Activities

Cas3 is a fast and highly processive molecular motor, which is recruited by Cascade through the PAM-dependent pathway and then translocates along the flanking DNA. This translocation does not coincide with any apparent DNA degradation or persistently unwound DNA. When Cas3 is recruited by Cascade through the PAM-dependent pathway, it always moves in the same direction along the DNA, consistent with expectations for $3' \rightarrow 5'$ translocation along the non-target strand. A subset of Cas3 molecules also forms optically detectable looped intermediates, and Cas3 likely generates smaller DNA loops that cannot be observed in

Formation of this longer-lived species prevents any further Cascade-specific recruitment of Cas3, most likely because the first stably bound Cas3 cleaves the R-loop, which destroys the Cas3 binding site (Mulepati and Bailey, 2013). Consistent with this interpretation, formation of stable Cascade-Cas3 intermediates coincides with the appearance of a ~200- to 300-nt ssDNA gap adjacent to the protospacer. The first molecule of Cas3 does not appear to induce any damage other than creating this initial ssDNA gap. This finding is notably different from bulk biochemical assays, which reveal more extensive DNA degradation (Mulepati and Bailey, 2013; Sinkunas et al., 2013). This difference may be explained by the potential for recruitment of additional Cas3 molecules in the bulk biochemical assays through a Cascade-independent pathway, as previously suggested (Mulepati and Bailey, 2013). Consistent with this explanation, Cas3 is a potent ssDNA nuclease even in the absence of Cascade (Mulepati and Bailey, 2013; Sinkunas et al., 2011, 2013). Thus, the ssDNA gaps generated by the first molecule of Cas3 likely reflect an early intermediate in the degradation pathway and serve as an entryway for additional ssDNA-specific nucleases, including Cas3 or perhaps other host enzymes. Together, these findings

our experiments, suggesting these looped intermediates may be a common feature of the PAM-dependent pathway (Figure 7B). Interestingly, similar looping behaviors have been reported for many different SF1 and SF2 helicases, including *Saccharomyces cerevisiae* Pif1 (Zhou et al., 2014), *Bacillus stearothermophilus* PcrA (Park et al., 2010), *E. coli* Rep (Myong et al., 2005), and *S. cerevisiae* Srs2 (Qiu et al., 2013). The looping behaviors exhibited by these proteins are thought to help establish and maintain a particular structural state of the DNA; for instance, PcrA and Srs2 repeatedly shuttle back and forth, while removing proteins from ssDNA proximal to an ssDNA/dsDNA junction to prevent aberrant recombination (Park et al., 2010; Qiu et al., 2013). Similarly, Pif1 repeatedly unwinds G-quadraplexes, ensuring that these structures do not inhibit DNA replication (Zhou et al., 2014). The looping activity observed for Cas3 may reflect attempts to dissociate from Cascade. Alternatively, looping may help keep the ssDNA gap clear of proteins, free of secondary structures or both, until the arrival of additional Cas3 molecules or other accessory nucleases.

PAM-Independent Target Recognition

Like the PAM-dependent search, the PAM-independent pathway also occurs by microscopic 3D diffusion, suggesting that Cascade must test for complementarity to the crRNA by either transiently melting the DNA or by taking advantage of the intrinsic breathing of the DNA duplex (Figure 7A). One primary difference between PAM-dependent and PAM-independent target recognition is that the efficiency of the PAM-independent pathway is comprised, such that a higher concentration of Cascade is required to achieve similar levels of occupancy at both targets. Despite this disparity in apparent association constants, Cascade can still bind tightly to DNA regardless of whether or not the protospacer has a canonical PAM. In both instances, the lifetime of the target-bound Cascade complexes is significantly longer that the typical doubling time of *E. coli*, a finding that is in good agreement with the results of magnetic tweezer experiments (Szczelkun et al., 2014). This tight binding would help ensure that even though escape target recognition is inefficient, in the rare instances in which an escape target is captured, Cascade would remain in place long enough to initiate downstream steps necessary for primed sequence acquisition (Figure 7C; see below). Interestingly, not all PAM mutations are equal with respect to Cascade, and the defect in binding with the ATT mutant PAM is more moderate that some other PAM mutations (Szczelkun et al., 2014). Future studies will be essential for testing the effects of other PAM mutations on target binding in these single-molecule assays.

Interestingly, recent single-molecule fluorescence resonance energy transfer (FRET) experiments have suggested that Cascade recognizes escape targets with substantially reduced fidelity, and interactions with these targets are characterized by a ~25-s lifetime (Blosser et al., 2015), which is identical to one of the nonspecific lifetimes observed in our experiments (Figure S1). We suggest that these shorter-lived complexes found by FRET reflect intermediates that have failed to transition into the more tightly bound complexes observed in our assays.

Importantly, PAM escape mutations reflect only a subset of mutations that can lead to a priming response, with the remainder occurring within the protospacer, but both types of escape mutants lead to similar priming responses (Datsenko et al., 2012; Fineran et al., 2014). We anticipate that Cascade will locate protospacer escape mutants through the normal PAM-dependent search pathway, but then may require Cas1-Cas2 to recruit Cas3 and initiate a priming response from this class of escape mutations.

Cas1-Cas2 Recruitment of Cas3 to Escape Targets

We demonstrate that the Cas1-Cas2 complex serves as a *trans*-acting factor necessary for the recruitment and regulation of Cas3 at protospacers bearing an escape PAM (Figure 7C). Recruitment may occur through one of two general mechanisms. Cas1-Cas2 may modify the structure of Cascade such that it can now directly recruit Cas3 by the same process as occurs during PAM-dependent recruitment. Alternatively, protein-protein contacts with Cas1-Cas2 may directly recruit Cas3 to the escape target through a mechanism that is distinct from the Cascade-dependent recruitment at cognate protospacers. Importantly, the behavior of Cas3 at the escape targets differs markedly from the behavior of Cas3 at cognate targets. First, Cas3 can translocate in either direction from the escape targets, implying that that Cas3 is loaded onto the flanking phage DNA through a different pathway than is observed at cognate protospacers. Second, there was no evidence that Cas3 generates ssDNA gaps at the escape targets, nor was there any evidence that Cas3 even nicked the DNA when loaded at escape targets, suggesting that the nuclease activity of Cas3 is fully attenuated at escape targets. The inability of Cas3 to cleave the escape target is also consistent with the fact that the vast majority of cells will die when infected with phage bearing an escape mutation, and immunity is only conferred for those rare survivors that successfully update the CRISPR locus (Datsenko et al., 2012). Third, Cas3 loaded at escape targets exhibited only a ~6-s lifetime prior to initiating translocation, but there was no evidence for the longer-lived intermediate (τ_2) that we have ascribed to ssDNA degradation. Fourth, there was no evidence for DNA looping when Cas3 initiated translocation from the escape target, suggesting that Cas3 is more readily released from Cascade at the escape target. Together, these observations suggest that Cas1-Cas2 recruits and loads Cas3 onto the DNA flanking the escape targets through a mechanism that is distinct from the Cascade-mediated mechanism that takes place at cognate protospacers.

Primed Acquisition of New Spacer Sequences

Together, our data provide direct support for a model of primed sequence acquisition involving Cas1-Cas2-mediated recruitment of Cas3 to Cascade at escape targets, followed by ATP-dependent translocation of Cas3 along the foreign DNA (Figure 7C). Cas3 can move in either direction away from the escape target, consistent with the expectation that new spacers can be acquired from either side of an escape target (Richter et al., 2014). Translocation of Cas3 away from the escape target does not induce DNA damage, and we speculate that Cas3 may be looking for an as-yet-unidentified signal (e.g., DNA sequence, partner protein, or both) necessary to activate its nuclease activity, or the nuclease activity of a partner protein, at some distal

location. Importantly, although the tagged Cas6e subunits remain bound to the protospacer after Cas3 translocation, we do not know whether the other Cas proteins are also left behind. It is possible that Cas3 takes a subset of Cascade components while translocating along the DNA. In fact, Cas3 is naturally linked with Cse1 in a single polypeptide chain in some systems, suggesting that Cse1 may have additional downstream functions during Cas3 translocation (Westra et al., 2012b). In addition, Cas1-Cas2 are essential to process and insert new spacer sequences into the CRISPR locus (Nuñez et al., 2014; Nuñez et al., 2015), and one attractive model is that Cas1-Cas2 travel with Cas3 as part of a larger spacer acquisition complex (Figure 7C), which would allow delivery of Cas1-Cas2 to sites distal to an escape target, where they would then be able to process the DNA to promote new spacer acquisition. In support of this model, studies in the closely related type 1F CRISPR-Cas system from *Pectobacterium atrosepticum* have shown that Cas3 interacts directly with Cas1 (Richter et al., 2012).

Early models suggested Cascade might diffuse away from the escape target (Datsenko et al., 2012). However, this model was later disfavored because the distribution of new spacers acquired from a circular plasmid was inconsistent with expectations for a diffusion-based mechanism, which would predict a strong bias toward acquisition of new spacer sequences near the original protospacer (Savitskaya et al., 2013). The high processivity of *E. coli* Cas3 (~12-kbp) explains why assays using relatively small plasmids (~5-kb) fail to yield a biased distribution of newly acquired spacer sequences as predicted by the original sliding hypothesis (Heler et al., 2014; Savitskaya et al., 2013). Interestingly, the type 1F CRISPR-Cas system from *P. atrosepticum* does exhibit a biased distribution of newly acquired spacers in response to an escape mutation (Richter et al., 2014). Assuming that priming occurs through a similar mechanism for the type 1F and type 1E CRISPR-Cas systems, our model predicts that *P. atrosepticum* Cas3 is less processive that *E. coli* Cas3, explaining why spacer acquisition bias can be observed in plasmid assays for *P. atrosepticum*.

Our data demonstrate that the first Cas3 molecule recruited to cognate protospacers through the PAM-dependent pathway can translocate rapidly away from Cascade before the DNA is destroyed. Moreover, the nuclease activity of Cas3 was partially attenuated by Cas1-Cas2 at cognate protospacers, allowing iterative Cas3 firing events presumably before the eventual destruction of the R-loop. Together, these observations suggest that priming might take place even when there is no escape mutation present in the invading DNA (Figure 7B). The ability to occasionally acquire new spacers in the absence of an escape mutation may allow microbes to routinely update the CRISPR locus even before foreign genetic elements have the opportunity to evade the CRIPSR/Cas immune response by acquiring new mutations.

EXPERIMENTAL PROCEDURES

Single-Molecule Assays

DNA curtains were fabricated by electron-beam lithography as previously described (Greene et al., 2010; Sternberg et al., 2014). A lipid bilayer was then deposited on the surface of the sample chamber; the anchor points were coated with anti-digoxigenin antibodies; and the DNA was anchored to the bilayer through a biotin-streptavidin linkage. The DNA was then aligned along the leading edges of the Cr diffusion barriers and coupled to the antibody-coated anchors through the application of hydrodynamic force. Cascade single-molecule binding assays were conducted in reaction buffer containing 40 mM Tris-HCl (pH 7.4), 1 mM MgCl$_2$, 25 mM KCl, 1 mg/ml BSA, 0.8% glucose, YOYO-1, and a glucose oxidase-catalase oxygen-scavenging system. The Cas6e subunit of Cascade was expressed with an N-terminal 3xFLAG tag, and the Cascade complex was labeled with antiFLAG-coated QDs (Invitrogen) for 10 min on ice prior to use. In experiments with Cas3, the YOYO-1 dye was omitted, and the reaction buffer was supplemented to contain 2 mM MgCl$_2$, 1 mM ATP, and 20 μM CoCl$_2$. Cas3 was labeled by incubation with streptavidin-coated QDs (Invitrogen) on ice for 20 min prior to injection onto the flow cell at 4 nM final concentration. RPA-eGFP labeling of ssDNA gaps was always performed at the end of the Cas3 experiments as a check for activity. In these assays, Cas3 was flushed from the sample chamber, followed by delivery of buffer containing 100 nM RPA-eGFP. The buffer flow was then terminated, and RPA-eGFP was incubated with the DNA for 10 min prior to imaging. Buffer conditions for experiments containing Cas1-Cas2 were identical to those above. Cas1 (8 nM) and Cas2 (16 nM) were pre-incubated on ice for 20 min and then mixed with Cas3 (4 nM) for an additional 5 min before being delivered to the flow cell. All single-molecule experiments were conducted at 25°C, unless otherwise indicated, and all data were collected and analyzed as previously described (Sternberg et al., 2014).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, one table, and four movies and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.003.

REFERENCES

Barrangou, R., and Marraffini, L.A. (2014). CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol. Cell *54*, 234–244.

Blosser, T.R., Loeff, L., Westra, E.R., Vlot, M., Künne, T., Sobota, M., Dekker, C., Brouns, S.J., and Joo, C. (2015). Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. Mol. Cell *58*, 60–70.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551–2561.

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. *3*, 945.

Fineran, P.C., Gerritzen, M.J., Suárez-Diez, M., Künne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H., and Brouns, S.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proc. Natl. Acad. Sci. USA *111*, E1629–E1638.

Gorman, J., Plys, A.J., Visnapuu, M.L., Alani, E., and Greene, E.C. (2010). Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. Nat. Struct. Mol. Biol. *17*, 932–938.

Gorman, J., Wang, F., Redding, S., Plys, A.J., Fazio, T., Wind, S., Alani, E.E., and Greene, E.C. (2012). Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. Proc. Natl. Acad. Sci. USA *109*, E3074–E3083.

Greene, E.C., Wind, S., Fazio, T., Gorman, J., and Visnapuu, M.L. (2010). DNA curtains for high-throughput single-molecule optical imaging. Methods Enzymol. *472*, 293–315.

Halford, S.E. (2009). An end to 40 years of mistakes in DNA-protein association kinetics? Biochem. Soc. Trans. *37*, 343–348.

Heler, R., Marraffini, L.A., and Bikard, D. (2014). Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. Mol. Microbiol. *93*, 1–9.

Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., and Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proc. Natl. Acad. Sci. USA *111*, 6618–6623.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J. Bacteriol. *169*, 5429–5433.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct *1*, 7.

Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. *60*, 174–182.

Mulepati, S., and Bailey, S. (2011). Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). J. Biol. Chem. *286*, 31896–31903.

Mulepati, S., and Bailey, S. (2013). In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. J. Biol. Chem. *288*, 22184–22192.

Myong, S., Rasnik, I., Joo, C., Lohman, T.M., and Ha, T. (2005). Repetitive shuttling of a motor protein on DNA. Nature *437*, 1321–1325.

Nuñez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., and Doudna, J.A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. Nat. Struct. Mol. Biol. *21*, 528–534.

Nuñez, J.K., Lee, A.S., Engelman, A., and Doudna, J.A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature *519*, 193–198.

Park, J., Myong, S., Niedziela-Majka, A., Lee, K.S., Yu, J., Lohman, T.M., and Ha, T. (2010). PcrA helicase dismantles RecA filaments by reeling in DNA in uniform steps. Cell *142*, 544–555.

Qiu, Y., Antony, E., Doganay, S., Koh, H.R., Lohman, T.M., and Myong, S. (2013). Srs2 prevents Rad51 filament formation by repetitive motion on DNA. Nat. Commun. *4*, 2281.

Richter, C., Gristwood, T., Clulow, J.S., and Fineran, P.C. (2012). In vivo protein interactions and complex formation in the Pectobacterium atrosepticum subtype I-F CRISPR/Cas System. PLoS ONE *7*, e49549.

Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H., and Fineran, P.C. (2014). Priming in the type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic Acids Res. *42*, 8516–8526.

Rollins, M.F., Schuman, J.T., Paulus, K., Bukhari, H.S., and Wiedenheft, B. (2015). Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. Nucleic Acids Res. *43*, 2216–2222.

Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksnys, V., and Seidel, R. (2015). Directional R-loop formation by the CRISPR-Cas surveillance complex cascade provides efficient off-target site rejection. Cell Rep., S2211-1247(15)00135-7.

Sashital, D.G., Wiedenheft, B., and Doudna, J.A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol. Cell *46*, 606–615.

Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A., and Severinov, K. (2013). High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in E. coli. RNA Biol. *10*, 716–725.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc. Natl. Acad. Sci. USA *108*, 10098–10103.

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO J. *30*, 1335–1342.

Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. EMBO J. *32*, 385–394.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62–67.

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc. Natl. Acad. Sci. USA *111*, 9798–9803.

van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat. Rev. Microbiol. *12*, 479–492.

von Hippel, P.H., and Berg, O.G. (1989). Facilitated target location in biological systems. J. Biol. Chem. *264*, 675–678.

Westra, E.R., Swarts, D.C., Staals, R.H., Jore, M.M., Brouns, S.J., and van der Oost, J. (2012a). The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. Annu. Rev. Genet. *46*, 311–339.

Westra, E.R., van Erp, P.B., Künne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K., et al. (2012b). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol. Cell *46*, 595–605.

Zhou, R., Zhang, J., Bochman, M.L., Zakian, V.A., and Ha, T. (2014). Periodic DNA patrolling underlies diverse functions of Pif1 on R-loops and G-rich DNA. eLife *3*, e02190.

Hypervulnerability to Sound Exposure through Impaired Adaptive Proliferation of Peroxisomes

Graphical Abstract

Before sound-exposure

$Pjvk^{+/+}$; $Pjvk^{-/-}$

cochlea   auditory hair cell

After sound-exposure

$Pjvk^{+/+}$ $Pjvk^{-/-}$

peroxisome proliferation   peroxisome degeneration

Authors

Sedigheh Delmaghani, Jean Defourny, Asadollah Aghaie, ..., Jean-Pierre Hardelin, Paul Avan, Christine Petit

Correspondence

christine.petit@pasteur.fr

In Brief

Hypervulnerability to sound exposure in pejvakin-deficient mice results from impaired adaptive proliferation of peroxisomes in response to the oxidative stress. In wild-type mice, loud sounds elicit adaptive peroxisomal proliferation.

Highlights

- Pejvakin-deficient mice and humans are hypervulnerable to sound exposure

- Oxidative stress induces a pejvakin-dependent proliferation of peroxisomes

- Peroxisome proliferation contributes to the physiological response to sound exposure

- *Pjvk* gene transfer can rescue auditory dysfunction in *Pjvk*$^{-/-}$ mice

Accession Number

GSE72722

CellPress

Hypervulnerability to Sound Exposure through Impaired Adaptive Proliferation of Peroxisomes

Sedigheh Delmaghani,[1,2,3] Jean Defourny,[1,2,3] Asadollah Aghaie,[2,3,4] Maryline Beurg,[5] Didier Dulon,[5] Nicolas Thelen,[6] Isabelle Perfettini,[1,2,3] Tibor Zelles,[7,8] Mate Aller,[7] Anaïs Meyer,[1,2,3] Alice Emptoz,[1,2,3] Fabrice Giraudet,[9,10,11] Michel Leibovici,[1,2,3] Sylvie Dartevelle,[12] Guillaume Soubigou,[13] Marc Thiry,[6] E. Sylvester Vizi,[7] Saaid Safieddine,[1,2,3] Jean-Pierre Hardelin,[1,2,3] Paul Avan,[9,10,11,15] and Christine Petit[1,2,3,4,14,15,*]

[1]Unité de Génétique et Physiologie de l'Audition, Institut Pasteur, 75015 Paris, France
[2]UMRS 1120, Institut National de la Santé et de la Recherche Médicale (INSERM), 75015 Paris, France
[3]Sorbonne Universités, UPMC Université Paris 06, Complexité du Vivant, 75005 Paris, France
[4]Syndrome de Usher et Autres Atteintes Rétino-Cochléaires, Institut de la Vision, 75012 Paris, France
[5]Equipe Neurophysiologie de la Synapse Auditive, Université de Bordeaux, Neurosciences Institute, CHU Pellegrin, 33076 Bordeaux, France
[6]Unit of Cell and Tissue Biology, GIGA-Neurosciences, University of Liege, CHU Sart-Tilman, B36, 4000 Liege, Belgium
[7]Institute of Experimental Medicine, Hungarian Academy of Sciences, 1083 Budapest, Hungary
[8]Department of Pharmacology and Pharmacotherapy, Semmelweis University, 1089 Budapest, Hungary
[9]Laboratoire de Biophysique Sensorielle, Université d'Auvergne, 63000 Clermont-Ferrand, France
[10]UMR 1107, Institut National de la Santé et de la Recherche Médicale (INSERM), 63000 Clermont-Ferrand, France
[11]Centre Jean Perrin, 63000 Clermont-Ferrand, France
[12]Plateforme d'Ingénierie des Anticorps, Institut Pasteur, 75015 Paris, France
[13]Plateforme Transcriptome et Épigénome, Institut Pasteur, 75015 Paris, France
[14]Collège de France, 75005 Paris, France
[15]Co-senior author
*Correspondence: christine.petit@pasteur.fr
http://dx.doi.org/10.1016/j.cell.2015.10.023

SUMMARY

A deficiency in pejvakin, a protein of unknown function, causes a strikingly heterogeneous form of human deafness. Pejvakin-deficient ($Pjvk^{-/-}$) mice also exhibit variable auditory phenotypes. Correlation between their hearing thresholds and the number of pups per cage suggest a possible harmful effect of pup vocalizations. Direct sound or electrical stimulation show that the cochlear sensory hair cells and auditory pathway neurons of $Pjvk^{-/-}$ mice and patients are exceptionally vulnerable to sound. Subcellular analysis revealed that pejvakin is associated with peroxisomes and required for their oxidative-stress-induced proliferation. $Pjvk^{-/-}$ cochleas display features of marked oxidative stress and impaired antioxidant defenses, and peroxisomes in $Pjvk^{-/-}$ hair cells show structural abnormalities after the onset of hearing. Noise exposure rapidly upregulates Pjvk cochlear transcription in wild-type mice and triggers peroxisome proliferation in hair cells and primary auditory neurons. Our results reveal that the antioxidant activity of peroxisomes protects the auditory system against noise-induced damage.

INTRODUCTION

Mutations of *PJVK*, which encodes pejvakin, a protein of unknown function present only in vertebrates, cause the DFNB59-recessive form of sensorineural hearing impairment. In the first patients described (Delmaghani et al., 2006), the impairment was restricted to neurons of the auditory pathway, with auditory brainstem responses (ABRs) displaying abnormally decreased wave amplitudes and increased inter-wave latencies (Starr and Rance, 2015). ABRs monitor the electrical response of auditory pathways to brief sound stimuli, from the primary auditory neurons synapsing with the sensory cells of the cochlea, the inner hair cells (IHCs), to the colliculus in the midbrain (Møller and Jannetta, 1983). However, some DFNB59 patients were found to have a cochlear dysfunction, as shown by an absence of the otoacoustic emissions (OAEs) that are produced by the outer hair cells (OHCs), frequency-tuned cells endowed with electro-motility that mechanically amplify the sound stimulation of neighboring IHCs (Ashmore, 2008). These patients had truncating mutations of *PJVK*, whereas the previously identified patients, with extant OAEs, had missense mutations (p.T54I or p.R183W) (Ebermann et al., 2007; Schwander et al., 2007; Borck et al., 2012). However, the identification of patients also carrying the p.R183W missense mutation but lacking OAEs (Collin et al., 2007) refuted any straightforward connection between the nature of the *PJVK* mutation and the hearing phenotype. The severity of deafness in DFNB59 patients varies from moderate to profound

Figure 1. Hearing Loss Variability and Greater Sensitivity to Controlled Sound Exposure in *Pjvk*$^{-/-}$ Mice

(A) ABR thresholds at 10 kHz in P30 *Pjvk*$^{+/+}$ (n = 26 mice) and *Pjvk*$^{-/-}$ (n = 48 mice) littermates.

(B) DPOAE thresholds at 10 kHz in P30 *Pjvk*$^{+/+}$ (n = 14 mice) and *Pjvk*$^{-/-}$ (n = 48 mice) littermates. In ears with no DPOAE, even at 75 dB SPL (the highest sound intensity tested), DPOAE thresholds were arbitrarily set at 80 dB SPL.

(C) Relationship between the number of pups raised together (determining sound levels in the immediate environment) and ABR thresholds at 10 kHz in P21 *Pjvk*$^{-/-}$ pups. Inset: a time-frequency analysis of a mouse pup's vocalization. Pup calls from P0 to P21 form harmonic series of about 5 kHz, with the most energetic harmonic at about 10 kHz. In a 12-pup litter, call levels reach 105 ± 5 dB SPL at the entrance to the ear canals of the pups.

(D) ABR thresholds at 10 kHz in P30 *Pjvk*$^{+/+}$ and *Pjvk*$^{-/-}$ mice before (dots) and after (crosses) controlled sound exposure. ns, not significant; ***p < 0.001. See also Figure S1.

and may even be progressive in some patients, suggesting that extrinsic factors may influence the hearing phenotype.

We investigated the role of pejvakin, with the aim of determining the origin of the phenotypic variability of the DFNB59 form of deafness. Our study of *Pjvk* knockout mouse models and of patients revealed an unprecedented hypervulnerability of auditory hair cells and neurons to sound exposure, accounting for phenotypic variability. We found that pejvakin is a peroxisome-associated protein involved in the oxidative-stress-induced proliferation of this organelle. Pejvakin-deficient mice revealed the key role of peroxisomes in the redox homeostasis of the auditory system and in the protection against noise-induced hearing loss.

RESULTS

Heterogeneity in the Hearing Sensitivity of *Pjvk*$^{-/-}$ Mice

We generated pejvakin-null (*Pjvk*$^{-/-}$) mice carrying a deletion of *Pjvk* exon 2, resulting in a frameshift at codon position 71

(p.Gly71*fs**9) (Figure S1; see the Supplemental Experimental Procedures). ABR thresholds recorded on postnatal day 30 (P30) *Pjvk*$^{-/-}$ mice (n = 48) ranged from 35 to 110 dB SPL (sound pressure level) at 10 kHz but never exceeded 30 dB SPL in their *Pjvk*$^{+/+}$ littermates (n = 26) (Figure 1A). This broad range of hearing sensitivity in *Pjvk*$^{-/-}$ mice, from near-normal hearing to almost complete deafness, extended across the whole frequency spectrum. The thresholds of distortion-product OAEs (DPOAEs) at 10 kHz (i.e., the minimum stimulus required for DPOAEs production by OHCs) also fell within an abnormally large range of values, from 30 to 75 dB SPL, in 28 *Pjvk*$^{-/-}$ mice, indicating an OHC dysfunction, and DPOAEs were undetectable in another 20 *Pjvk*$^{-/-}$ mice, suggesting a complete OHC defect (Figure 1B). The absence of pejvakin in mice thus results in a puzzlingly large degree of hearing phenotype variability.

Hypervulnerability to the Natural Acoustic Environment in *Pjvk*$^{-/-}$ Mice

We investigated the variability of *Pjvk*$^{-/-}$ auditory phenotypes, by first determining the ABR thresholds of *Pjvk*$^{-/-}$ littermates from different crosses. Large differences were observed between crosses, with much fewer differences between the *Pjvk*$^{-/-}$ littermates of individual crosses. Litters with larger numbers of pups (6 to 12) had higher ABR thresholds, suggesting that the natural acoustic environment, with the calls of larger numbers of pups, might be deleterious in *Pjvk*$^{-/-}$ mice. Pups are vocally very active from birth to about P20. We manipulated the level of exposure to pup calls by randomly splitting large litters of *Pjvk*$^{-/-}$ pups into groups of 2, 4, 6 and 10 pups per cage, with foster mothers, before P10, i.e., several days before hearing onset. The ABR thresholds at P21 were significantly correlated with the number of pups raised together (p < 0.001, r^2 = 0.51) (Figure 1C).

We then evaluated the effect of a controlled sound stimulation on hearing, by presenting 1,000 tone bursts at 10 kHz, 105 dB SPL (2-ms plateau stimulations separated by 60-ms intervals of silence), energetically equivalent to a 3-min stay in the natural environment of a 12-pup litter, while monitoring the ABRs during sound exposure. These conditions are referred to hereafter as "controlled sound exposure." We probed the effect of sound exposure by ABR tests, which, limited to 50 repetitions of tone bursts, did not influence the hearing thresholds of *Pjvk*$^{-/-}$ mice. In a sample of P30 *Pjvk*$^{-/-}$ mice with initial ABR threshold elevation (below 35 dB SPL), controlled sound exposure affected ABR thresholds in the 12–20 kHz frequency interval (corresponding to the cochlear zones in which hair-cell stimulation was strongest), with an immediate increase of 21.7 ± 10.3 dB (n = 8; p < 0.001), not observed in *Pjvk*$^{+/+}$ mice (2.2 ± 2.4 dB, n = 12; p = 0.3) (Figure 1D). *Pjvk*$^{-/-}$ mice transferred to a silent environment after exposure displayed a further increase of 33.7 ± 16.0 dB (n = 8) 2 days after exposure. The threshold shift decreased to 23.7 ± 18.0 dB at 7 days, and disappeared entirely by 14 days. When exposed mice were returned to the box with their littermates, their ABR continued to increase, at a rate of 15 dB per week. Pejvakin deficiency thus results in particularly high levels of vulnerability to low levels of acoustic energy, and the increase in ABR thresholds is reversible but only slowly and in a quiet environment.

Hair Cells and Auditory Pathway Neurons Are Affected by Pejvakin Deficiency

To identify the cellular targets of the pejvakin deficiency, we specifically probed the function of auditory hair cells and neurons in $Pjvk^{-/-}$, hair cell-conditional $Pjvk$ knockout ($Pjvk^{fl/fl}$Myo15-$cre^{+/-}$), and $Pjvk^{+/+}$ mice, at the age of 3 weeks, before and after controlled sound exposure or controlled electrical stimulation. The responses of the IHCs to sound-induced vibrations amplified by OHCs trigger action potentials in the distal part of primary auditory neurons, at the origin of ABR wave I. In $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice, which lack pejvakin only in the hair cells, ABR wave I amplitude and latency at 105 dB SPL specifically probed IHC function, because IHC responses to such loud sounds are independent of OHC activity (Robles and Ruggero, 2001). The larger wave I latency (1.58 ms in $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice [n = 20] versus 1.32 ms in $Pjvk^{+/+}$ littermates [n = 30]; p < 0.001) and lower wave I amplitude (37% of the amplitude in $Pjvk^{+/+}$ littermates; p < 0.001) suggested a dysfunction of the IHCs. Controlled sound exposure induced further decreases in ABR wave I amplitude in $Pjvk^{-/-}$ and $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice (48% and 55% of pre-exposure amplitude, respectively) with respect to $Pjvk^{+/+}$ mice (108%; p < 0.001 for both comparisons) (Figure 2A), demonstrating that $Pjvk^{-/-}$ IHCs are hypervulnerable to sound. As shown above, OHCs are also affected by the pejvakin deficiency. Controlled sound exposure triggered a mean decrease in the DPOAE amplitude of 16.9 ± 7.2 dB in the 12 to 20 kHz frequency interval in $Pjvk^{-/-}$ mice with persistent DPOAEs (n = 8; p < 0.0001), and an increase in DPOAE threshold, but it had no effect on the DPOAEs of $Pjvk^{+/+}$ mice (n = 9; p = 0.51) (Figure 2B). OHCs lacking pejvakin are thus also hypervulnerable to sound.

We investigated the effect of the absence of pejvakin on the auditory pathway by comparing electrically evoked brainstem responses (EEBR) in $Pjvk^{-/-}$ and $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice (see the Supplemental Experimental Procedures). The amplitudes of the most distinctive EEBR waves, E II and E IV, did not differ between the two types of mice (for wave E IV: 2.6 ± 1.8 μV in $Pjvk^{-/-}$ mice [n = 18] and 2.2 ± 1.2 μV in $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice [n = 11]; t test, p = 0.13). However, following controlled electrical exposure at 200 impulses/s for 1 min, as opposed to electric-impulse stimulation with 16 impulses/s for 10 s for pre- and post-exposure EEBR tests, E II and E IV EEBR wave amplitudes got 41% and 47% smaller, respectively, for at least 3 min, in $Pjvk^{-/-}$ mice (n = 5; paired t test, p = 0.02 and p = 0.01, respectively), but were unaffected in $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice (n = 10; p = 0.83) (Figures 2D and 2G–2I). The E II-E IV interwave interval was 0.41 ms longer in $Pjvk^{-/-}$ mice (n = 5) than in $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice (n = 10; p = 0.003), and controlled electrical exposure extended this interval by a further 0.15 ms in $Pjvk^{-/-}$ mice only (paired t test, p = 0.001) (Figures 2H and 2I). Likewise, the latency interval between ABR wave I and wave IV (the counterpart of wave E IV), abnormal in one-third of the $Pjvk^{-/-}$ mice tested (with an ABR threshold < 95 dB SPL, n = 12) (Figures 2C and 2E), got abnormal in all of them after controlled sound exposure (0.16 ms further increase; paired t test, p < 0.001). By contrast, it remained normal in $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice (n = 10 ears; p = 0.73) (Figures 2C and 2F). Thus, the absence of pejvakin affects the propagation

of action potentials in the auditory pathway after both controlled electrical and sound exposure in the $Pjvk^{-/-}$ mice.

To clarify whether these abnormalities were of neuronal or glial origin, we performed a rescue experiment in $Pjvk^{-/-}$ mice, using adeno-associated virus 8 (AAV8) vector-mediated transfer of the murine pejvakin cDNA (AAV8-Pjvk). AAV8 injected into the cochlea transduces the primary auditory neurons (cochlear ganglion neurons) and neurons of the cochlear nucleus (Figure S2A), but not the hair cells. All $Pjvk^{-/-}$ mice (n = 7) injected on P3 and tested on P21 had normal ABR interwave I-IV latencies (Figure 2J), and their EEBR wave-E IV amplitude was insensitive to controlled electrical stimulation (1.91 ± 0.97 μV before and 1.87 μV ± 1.07 after stimulation; paired t test, p = 0.59) (Figures 2K and 2L). The absence of pejvakin thus renders auditory pathway neurons hypervulnerable to exposure to mild, short sound stimuli.

Hypervulnerability to Sound in DFNB59 Patients

We then investigated whether the hearing of DFNB59 patients was also hypervulnerable to sound exposure. We tested five patients carrying the p.T54I mutation (Delmaghani et al., 2006). Transient-evoked OAEs (TEOAEs) assessing OHC function over a broad range of frequencies were detected for all ears, despite the severe hearing impairment (hearing threshold increasing from 66 dB HL at 250 Hz to 84 dB at 8 kHz). Following minimal exposure to impulse stimuli (clicks at 99 dB nHL), ABR waves were clearly identified in response to 250 clicks. When exposure was prolonged to 1,000 clicks (the standard procedure), wave V, the equivalent of mouse ABR-wave IV, which was initially conspicuous, displayed a decrease in amplitude (to 39% ± 30% of its initial amplitude) and an increase in latency (of 0.30 ± 0.15 ms) (Figures 3A, 3C, and 3D). In parallel, the I-V interwave interval increased by 0.30 ± 0.15 ms. Wave-V amplitude and latency recovered fully after 10 min of silence (Figure 3B). In control patients with sensorineural hearing impairment of cochlear origin matched for ABR thresholds, similar sound stimulation did not affect ABR wave-V amplitude (105% ± 14% of the initial amplitude after exposure; n = 13 patients) or latency (−0.02 ± 0.07 ms change after exposure) (Figures 3C and 3D). Exposure of the DFNB59 patients to 1,000 clicks also affected TEOAEs (6.1 ± 5.2 dB nHL decrease in amplitude; paired t test, p = 0.02). Therefore, as in pejvakin-deficient mice, the cochlear and neuronal responses of DFNB59 patients were affected by exposure to low-energy sound.

Redox Status Abnormalities and ROS-Induced Cell Damage in the Cochlea of $Pjvk^{-/-}$ Mice

We studied the $Pjvk^{-/-}$ cochlea by light microscopy on semithin sections and electron microscopy. On P15 and P21, both OHCs and IHCs were normal in number and shape. Their hair bundles (the mechanoreceptive structures responding to sound), the ribbon synapses of the IHCs, and their primary auditory neurons were unmodified (data not shown). On P30, we observed the loss of a few OHCs (16% ± 11%, n = 5 mice), restricted to the basal region of the cochlea (tuned to high-frequency sounds). From P30 onward, OHCs, cochlear ganglion neurons, and then IHCs disappeared, and the sensory epithelium (organ of Corti) progressively degenerated (Figure S3A).

Figure 2. Effects on Auditory Function of Brief Exposure to Moderately Intense Stimuli in *Pjvk*[+/+], *Pjvk*[−/−], and *Pjvk*[fl/fl]Myo15-*cre*[+/−] Mice

(A–C) ABR wave I amplitude (A), DPOAE amplitude (B), and ABR interwave I-IV latency (C) in *Pjvk*[+/+], *Pjvk*[−/−], and *Pjvk*[fl/fl]Myo15-*cre*[+/−] mice, before (dots) and after (crosses) controlled sound exposure, revealing the hypervulnerability to sound of both types of cochlear hair cells (IHCs and OHCs) and of the neural pathway.

(D) EEBR wave E IV amplitude before and after controlled electrical exposure in *Pjvk*[−/−] and *Pjvk*[fl/fl]Myo15-*cre*[+/−] mice was abnormal and hypervulnerable only when pejvakin is absent from auditory neurons (*Pjvk*[−/−] mice).

(E and F) Examples of ABRs in *Pjvk*[−/−] and *Pjvk*[fl/fl]Myo15-*cre*[+/−] mice: the latency of wave I is affected by controlled sound exposure in both mutant mice, and wave IV displays an additional increase in latency only in *Pjvk*[−/−] mice.

(G–I) Examples of EEBRs in *Pjvk*[+/+] (G), *Pjvk*[−/−] (H), and *Pjvk*[fl/fl]Myo15-*cre*[+/−] (I) mice; EEBRs are affected by controlled electrical exposure only in *Pjvk*[−/−] mice.

(legend continued on next page)

We investigated possible changes in gene expression in the organ of Corti of P15 $Pjvk^{-/-}$ mice, by microarrays (see the Supplemental Experimental Procedures). Eighteen genes had expression levels at least 1.5-fold higher or lower in $Pjvk^{-/-}$ mice than in $Pjvk^{+/+}$ mice. Marked differences were observed for four genes involved in the redox balance—$CypA$, $Gpx2$, c-Dct, and $Mpv17$—encoding cyclophilin A, glutathione peroxidase 2, c-dopachrome tautomerase, and Mpv17, respectively (Table S1). All of these genes were downregulated in $Pjvk^{-/-}$ mice, a result confirmed by qRT-PCR (Figure S4A), and all encode antioxidant proteins, suggesting that $Pjvk^{-/-}$ mice have impaired antioxidant defenses (Table S1).

We thus assessed the level of oxidative stress in the cochlea of P21 $Pjvk^{-/-}$ mice, by determining the ratio of reduced to oxidized glutathione (GSH:GSSG). The GSSG content was about three times larger than in $Pjvk^{+/+}$ mice, whereas the GSH content was 23% smaller, resulting in a GSH:GSSG ratio in $Pjvk^{-/-}$ cochleas reduced by a factor of 3.4 (Figure 4A). Pejvakin deficiency thus results in cochlear oxidative stress.

We assessed lipid peroxidation by reactive oxygen species (ROS) in $Pjvk^{-/-}$ mice, by immunofluorescence-based detection of the by-product 4-hydroxy-2-nonenal (4-HNE). Strong immunoreactivity was observed in P60 $Pjvk^{-/-}$ hair cells and cochlear ganglion neurons (Figure S3B). Quantification of lipid peroxidation in microdissected organs of Corti from P30 $Pjvk^{-/-}$ and $Pjvk^{+/+}$ mice, showed a moderate, but statistically significant, increase of the malondialdehyde content in the absence of pejvakin (2.15 ± 0.14 μM in $Pjvk^{-/-}$ versus 1.84 ± 0.11 μM in $Pjvk^{+/+}$ mice; p = 0.04). Thus, pejvakin deficiency led to impaired antioxidant defenses in the cochlea, resulting in ROS-induced cell damage.

We then studied electrophysiological features of IHCs and OHCs in the mature cochlea of P19–P21 $Pjvk^{-/-}$ mice. In IHCs, the number of synaptic ribbons, Ca^{2+} currents, and synaptic exocytosis were unaffected (Figure S5A). We investigated whether $Pjvk^{-/-}$ mice display the main K^+ currents found in mature IHCs, specifically $I_{K,f}$, which plays a major part in IHC repolarization and is involved in the high temporal precision of action potentials in postsynaptic nerve fibers, $I_{K,s}$, and $I_{K,n}$ (Oliver et al., 2006). The $I_{K,f}$ current that flows through the large conductance voltage- and Ca^{2+}-activated potassium (BK) channels, a well-known target of ROS (Tang et al., 2004), was detected in only 4 out of 11 $Pjvk^{-/-}$ IHCs, and the mean number of spots immunolabeled for the BK α-subunit per IHC was much lower in $Pjvk^{-/-}$ mice (5.0 ± 1.4, n = 283 IHCs from seven mice) than in $Pjvk^{+/+}$ mice (13.9 ± 2.6, n = 204 IHCs from nine mice; t test, p < 0.001). By contrast, the $I_{K,s}$ and $I_{K,n}$ currents were not affected (Figures 4B and S5B). The electromotility of OHCs was moderately impaired in $Pjvk^{-/-}$ mice (Figure S5C). This contrasted with the total loss of DPOAE in a large majority of Pjvk$^{-/-}$ mice from P15 on, even at the highest possible stimulus level of 75 dB SPL. It thus pinpointed the existence of an additional defect, likely a mechanoelectrical transduction defect, the

main determinant of DPOAEs at high stimulus levels (Avan et al., 2013). The decrease of the cochlear microphonic potential that reflects mechanoelectrical transduction currents through OHCs of the basal-most cochlear region, indeed corroborated the DPOAE measurements: this potential, recorded for a 5-kHz sound stimulus at 95 dB SPL, was always larger than 10 μV in $Pjvk^{+/+}$ mice (n = 8), but fell between 5 and 3 μV in the P21 $Pjvk^{-/-}$ mice with residual DPOAEs (n = 2), and below 1 μV, in the $Pjvk^{-/-}$ mice without persisting DPOAEs (n = 6). Taken together, oxidative stress in the $Pjvk^{-/-}$ cochlea impacts various electrophysiological properties of the hair cells, particularly mechanoelectrical transduction and K^+ current through BK channels.

Mitochondrial defects are a common cause of ROS overproduction. However, we did not find evidence that mitochondria were damaged, as vulnerability of the mitochondrial membrane potential, $\Delta\psi_m$, to the uncoupler carbonyl cyanide 4-(trifluoromethoxy)phenylhydrazone (FCCP) in the organ of Corti and cochlear ganglion was similar in P17-P30 $Pjvk^{-/-}$ and $Pjvk^{+/+}$ mice, and analysis of $Pjvk^{-/-}$ hair cells by transmission electron microscopy (TEM) revealed no mitochondrial abnormalities (Figure S5D; data not shown).

Pejvakin Is a Peroxisome-Associated Protein

By using $Pjvk^{-/-}$ cochlea as control, we found that neither the commercially available antibodies nor our initial polyclonal antibody (Delmaghani et al., 2006) specifically recognized pejvakin (data not shown). Given the limited divergence of the pejvakin amino-acid sequence among vertebrates, we tried to elicit an immune response in $Pjvk^{-/-}$ mice (see the Experimental Procedures). The monoclonal antibody obtained, Pjvk-G21, labeled peroxisomes stained by peroxisome membrane protein 70 (PMP70) antibodies in transfected HeLa cells expressing pejvakin (Figure S6A) and in the human HepG2 hepatoblastoma cell line, which is particularly rich in this organelle (Figure 5A). The specificity of the Pjvk-G21 antibody was demonstrated by the immunolabeling of peroxisomes in the hair cells of $Pjvk^{+/+}$, but not of $Pjvk^{-/-}$ and $Pjvk^{fl/fl}$Myo15-cre$^{+/-}$ mice (Figures 5B and S6B).

Prediction programs failed to detect the PTS1 or PTS2 motifs in the pejvakin sequence (Mizuno et al., 2008), the targeting signals for the importation of peroxisomal matrix proteins into the organelle (Smith and Aitchison, 2013), suggesting that pejvakin is a peroxisomal membrane or membrane-associated protein.

Structural Abnormalities of Peroxisomes in the Hair Cells of $Pjvk^{-/-}$ Mice

We investigated the distribution and morphology of peroxisomes by TEM. Peroxisomes were identified on the basis of catalase activity detection using 3,3′-diaminobenzidine as substrate. We focused on OHCs, the first to display a dysfunction in $Pjvk^{-/-}$ mice. On P30, but not on P15, both the distribution and shape of peroxisomes differed between $Pjvk^{-/-}$ and $Pjvk^{+/+}$ mice (Figure 5E). In $Pjvk^{+/+}$ OHCs, the peroxisomes were restricted to an area immediately below the cuticular plate. In $Pjvk^{-/-}$ mice,

---

(J–L) Neuronal function rescue in $Pjvk^{-/-}$ mice by transduction with AAV8-Pjvk: effects on ABR interwave I-IV latency (J), on EEBR wave E IV amplitude and its hypervulnerability to electrical stimulation (K), and on EEBR interwave E II-E IV latency (one example is shown in L, to be compared with H). Vertical arrows indicate the positions of waves I and IV on ABR traces and of waves E II and E IV on EEBR traces. ns, not significant; ***p < 0.001. Error bars represent the SD. See also Figures S1 and S2A.

A First sound exposure

B Second sound exposure after 10 min in silence

C **D**

Figure 3. Hypervulnerability to Sound in DFNB59 Patients

(A) ABR waves I, III, and V (vertical arrows) in one ear of a patient carrying the *PJVK* p.T54I mutation, in response to 250, 500, and 1000 impulse stimuli (clicks) at 99 dB nHL.

(B) Repeated ABRs after 10 min of silence, with an even larger vulnerability of waves I, III, and V.

(C and D) Distributions of the amplitude (C) and latency (D) of ABR wave V in the tested sample of p.T54I patients (n = 8 ears), and in a control group of patients

the peroxisomes located just below the cuticular plate were slightly larger than those in $Pjvk^{+/+}$ mice. Strikingly, irregular catalase-containing structures, some of which were juxtaposed, were present in the perinuclear region, in the immediate vicinity of the nuclear membrane of all $Pjvk^{-/-}$ OHCs, but not of $Pjvk^{+/+}$ OHCs (Figure 5E). The lack of pejvakin thus results in peroxisome abnormalities in OHCs after the onset of hearing.

Pejvakin Is Involved in Oxidative Stress-Induced Peroxisome Proliferation

In HepG2 cells, protrusions emerging from some peroxisomes, the first step of peroxisome biogenesis from pre-existing peroxisomes, were immunoreactive for pejvakin. String-of-beads structures corresponding to elongated and constricted peroxisomes, preceding final fission (Smith and Aitchison, 2013), were also pejvakin-immunoreactive, suggesting a role for this protein in peroxisome proliferation (Figure S6C). Peroxisomes actively contribute to cellular redox balance, by producing and scavenging/degrading H_2O_2 through a broad spectrum of oxidases and peroxidases (especially catalase), respectively (Schrader and Fahimi, 2006). Because $Pjvk^{-/-}$ mice displayed features of marked oxidative stress in the cochlea, we investigated the possible role of pejvakin in peroxisome proliferation in response to oxidative stress induced by H_2O_2 (Lopez-Huertas et al., 2000). Embryonic fibroblasts derived from $Pjvk^{+/+}$ and $Pjvk^{-/-}$ mice were exposed to H_2O_2 (see the Supplemental Experimental Procedures). In unexposed cells, the number of peroxisomes was similar between the two genotypes (t test, p = 0.82). After H_2O_2 treatment, it increased by 46% in $Pjvk^{+/+}$ fibroblasts (p = 0.004), but remained unchanged in $Pjvk^{-/-}$ fibroblasts (p = 0.83), resulting in a statistically significant difference between the two genotypes (p < 0.001) (Figures 5C and S7A).

We then asked whether mutations reported in DFNB59 patients also affect peroxisome proliferation. We assessed the number of peroxisomes in transfected HeLa cells producing EGFP alone, EGFP and murine pejvakin, or EGFP and one of the mutated forms of murine pejvakin carrying the mutations responsible for DFNB59 (p.T54I, p.R183W, p.C343S, or p.V330Lfs*7). Cells producing the non-mutated pejvakin had larger numbers of peroxisomes than cells producing EGFP alone, whereas cells producing any of the mutated forms of pejvakin (mutPjvk-IRES-EGFP) had smaller peroxisome numbers. In addition, many of these cells contained enlarged peroxisomes, a feature typical of peroxisome proliferation disorders (Ebberink et al., 2012) (Figure 5D and S7B). Together, these results strongly suggest that pejvakin is directly involved in the production of new peroxisomes from pre-existing peroxisomes.

Upregulation of *Pjvk* Cochlear Transcription and Peroxisome Proliferation in Response to Sound

We then asked whether pejvakin is involved in the physiological response to sound. We first assessed the transcription of *Pjvk*

(n = 13) with cochlear hearing impairment and matched ABR thresholds, before and after exposure to clicks #250 to #1000. Boxes extend from the 25th to the 75th percentile. Horizontal bars and vertical bars indicate median values and extremes, respectively. Unlike the unaffected controls, all p.T54I patients displayed markedly decreased amplitudes and increased latencies.

A

B

F-actin / BK α-subunit / DAPI

and of *CypA*, *Gpx2*, *c-Dct*, and *Mpv17*, which were downregulated in *Pjvk*⁻/⁻ mice, in microdissected organs of Corti from P21 wild-type mice, with or without prior sound stimulation (5–20 kHz, 105 dB SPL for 1 hr; see the Supplemental Experimental Procedures). Transcript levels were analyzed by qRT-PCR at various times (1, 3, 6, and 18 hr) after sound exposure (Figure 6A). *Pjvk* transcript levels had increased by factors of 1.9 ± 0.1 and 3.5 ± 0.7, mean ± SEM, after 1 and 6 hr, respectively. *CypA*, *c-Dct*, and *Mpv17* were also upregulated after 6 hr (by factors of 6.6 ± 1.2, 4.3 ± 0.6, and 1.5 ± 0.1, respectively), as were *c-Fos* and *Hsp70*, used as a positive control, but not *Gpx2*. Thus, noise exposure leads to an upregulation of the transcription of *Pjvk* and of genes downregulated in *Pjvk*⁻/⁻ mice, and this effect is dependent on acoustic energy level of the stimulation (Figure S4B).

This result predicted that sound exposure would lead to peroxisome proliferation in the auditory system of wild-type mice. 6 hr after exposure (5–20 kHz, 105 dB SPL for 1 hr), the numbers of peroxisomes were unchanged (34.5 ± 0.8 and 35.9 ± 1.0, mean ± SEM, per IHC from unexposed and sound-exposed mice, respectively, n = 75 cells from six mice; t test, p = 0.25). However, at 48 hr, they had markedly increased, by a factor of 2.3, in both IHCs and OHCs (84.7 ± 5.0 per IHC and 16.5 ± 1.0 per OHC, n = 90 cells and n = 150 cells from six mice, respectively) compared to unexposed mice (36.8 ± 3.0 per IHC and 7.3 ± 0.4 per OHC, n = 90 cells and n = 150 cells from six mice, respectively; t test, p < 0.0001 for both comparisons). The number of peroxisomes had also increased, by 35%, in the dendrites of primary auditory neurons (1.7 ± 0.1 and 2.3 ± 0.2 peroxisomes per micrometer of neurite length, n = 40 neurites from five unexposed and five sound-exposed *Pjvk*⁺/⁺ mice, respectively; t test, p = 0.003) (Figure 6B).

Therapeutic Approaches in *Pjvk*⁻/⁻ Mice

Based on these results, we tested whether the classical antioxidant drug N-acetyl cysteine (NAC) (either alone or associated with α-lipoic acid and α-tocopherol; see the Supplemental Experimental Procedures) administered to *Pjvk*⁻/⁻ pups could improve their hearing. The ABR thresholds of P21 NAC-treated *Pjvk*⁻/⁻ pups (n = 21) were about 10 dB lower than those of untreated *Pjvk*⁻/⁻ pups (n = 24) for all frequencies tested (t test, p < 0.001 for all comparisons) (Figure 7A). The amplitude of the ABR wave I elicited at 105 dB SPL (4.35 ± 1.16 μV, n = 21) was the same as that of *Pjvk*⁺/⁺ mice (4.36 ± 1.15 μV, n = 18; t test, p = 0.97) and greater than that of untreated *Pjvk*⁻/⁻ mice (1.88 ± 1.07 μV, n = 24; t test, p < 0.001) (Figure 7B). EEBRs were more resistant to the high-rate electrical stimulation in treated than in untreated mutant mice (Figure 7C). Conversely, NAC had no beneficial effect on OHCs (data not shown). The association of NAC with α-lipoic acid and α-tocopherol did not perform any better (data not shown).

Full recovery of the neuronal phenotype was achieved by the intracochlear injection of AAV8-Pjvk (see above). As hair cells are not transduced by AAV8, we investigated whether AAV2/8, which transduces hair cells only (Figure S2B), could rescue the *Pjvk*⁻/⁻ hair-cell phenotype. The auditory function of *Pjvk*⁻/⁻ mice (n = 7, four pups per cage in every experiment) receiving intracochlear injections of AAV2/8-Pjvk-IRES-EGFP on P3 was assessed on P21, and the percentage of transduced IHCs and OHCs was evaluated in each injected and contralateral (not injected) cochlea, on the basis of EGFP fluorescence. Improvements in ABR thresholds of 20 to 30 dB SPL with respect to untreated mice were observed for frequencies between 10 and 20 kHz (t test, p < 0.001 for all comparisons; Figure 7D). Upon injection of AAV2/8-EGFP, DPOAEs, ABR thresholds, and ABR wave I amplitude and latency were similar to those of untreated *Pjvk*⁻/⁻ mice (data not shown). A partial reversion of the OHC dysfunction was obtained, with detectable DPOAEs in pejvakin cDNA-treated cochleas (threshold 54.0 ± 10.7 dB), but not in contralateral, untreated cochleas (Figure 7E). DPOAE thresholds were linearly correlated (r² = 0.74, p < 0.001) with the number of EGFP-tagged OHCs (Figure 7F), suggesting that the normalization of DPOAE thresholds may be possible if all OHCs could be transduced. The latency of the ABR wave I in response to a 105 dB SPL stimulation decreased significantly (1.38 ± 0.11 ms for the treated ears; n = 6, versus 1.53 ± 0.10 ms

Figure 5. Pejvakin Is a Peroxisome-Associated Protein Involved in the Oxidative Stress-Induced Peroxisomal Proliferation

(A and B) Immunolabeling of PMP70 and endogenous pejvakin in a HepG2 cell (A) and in two P20 $Pjvk^{+/+}$ IHCs (B). See also Figure S6B.

(C) Number of peroxisomes in $Pjvk^{+/+}$ and $Pjvk^{-/-}$ mouse embryonic fibroblasts (MEFs) subjected to 0.5 mM H_2O_2 versus untreated MEFs (n = 30 cells for each condition). See also Figure S7A.

(D) Untransfected HeLa cells (NT) and transfected cells producing either EGFP alone or EGFP, together with the wild-type pejvakin (Pjvk) or a mutated Pjvk (p.T54I, p.R183W, p.C343S, or p.V330Lfs*7). Left panel: bar chart showing the numbers of peroxisome per cell 48 hr after transfection. There were on average

(legend continued on next page)

Figure 6. Effect of Exposure to Loud Sounds on the Cochlear Expression of *Pjvk* and the Number of Peroxisomes in Cochlear Hair Cells and Ganglion Neurons

(A) *Pjvk*, *c-Dct*, *CypA*, *Mpv17*, and *Gpx2* transcript levels assessed by qRT-PCR in P21 *Pjvk*[+/+] organ of Corti 1, 3, 6, and 18 hr after sound exposure (5–20 kHz, 105 dB SPL for 1 hr). The levels of *c-Fos* and *Hsp70* transcripts were used as positive controls. See also Figure S4B.

(B) Peroxisome proliferation in P21 *Pjvk*[+/+] hair cells and cochlear ganglion neurons after sound exposure (same conditions as in A). Peroxisomes were counted 48 hr after sound exposure. OHCs, IHCs, and neuronal processes stained for F-actin, myosin VI, and neurofilament protein NF200, respectively. In OHCs and IHCs, the peroxisomes are located below the CP and throughout the cytoplasm, respectively. For OHCs, both a lateral view and a transverse optical section at the level of CP (scheme on the right) are shown. The number of peroxisomes was increased in OHCs, IHCs, and dendrites after sound exposure. N, cell nucleus. ***p < 0.001. Error bars represent the SEM. Scale bars are 5 μm.

See also Figure S4 and Table S1.

Before sound exposure, the numbers of peroxisomes in IHCs of P21 *Pjvk*[−/−] and AAV2/8-Pjvk-IRES-EGFP-injected *Pjvk*[−/−] mice did not differ from that of *Pjvk*[+/+] mice (30.5 ± 1.9, 32.3 ± 2.1, and 36.8 ± 3.0 peroxisomes, mean ± SEM per IHC, n = 60 cells from four *Pjvk*[−/−] and four AAV2/8-Pjvk *Pjvk*[−/−] mice, and n = 90 cells from six *Pjvk*[+/+] mice, respectively; t test, p = 0.11 and p = 0.30, respectively). By contrast, 48 hr after sound exposure (5–20 kHz) at 105 dB SPL for 1 hr, the number of peroxisomes had

for the contralateral, untreated ears; paired t test, p = 0.03 (Figure 7G), and its amplitude increased into the normal range (7.34 ± 0.80 μV versus 2.93 ± 0.92 μV; paired t test, p < 0.001) (Figure 7H), in relation to the number of EGFP-tagged IHCs (r^2 = 0.89 for wave I amplitude, p < 0.001; Figure 7I). No correction of the interwave I–IV latency was observed, as expected (data not shown).

Finally, we investigated the effect of the transduction of *Pjvk*[−/−] IHCs by AAV2/8-Pjvk-IRES-EGFP on their peroxisomes.

decreased by 63% in *Pjvk*[−/−] IHCs (30.5 ± 1.9 and 11.2 ± 1.3 peroxisomes per IHC, n = 75 cells from five unexposed and five sound-exposed *Pjvk*[−/−] mice, respectively; t test, p < 0.0001), and enlarged PMP70-labeled structures were present close to the nucleus (Figure 7J). In response to the same sound but of a lower intensity, i.e., 97 dB SPL for 1 hr, the number of peroxisomes was unchanged in *Pjvk*[−/−] IHCs (30.5 ± 1.9 and 34.6 ± 2.3 peroxisomes per IHC, n = 60 cells from four unexposed and four sound-exposed *Pjvk*[−/−] mice,

33% more peroxisomes in cells producing both EGFP and Pjvk (n = 200) than in cells producing EGFP alone (n = 150). Right panel: for every range of enlarged peroxisome size, x (0.6–0.8 μm, 0.8–1.0 μm, and >1.0 μm), in two perpendicular directions, the proportion of cells containing at least one peroxisome. See also Figure S7B.

(E) Abnormalities in shape and distribution of peroxisomes in mature *Pjvk*[−/−] OHCs detected by TEM (P30 *Pjvk*[−/−] [middle and right] and *Pjvk*[+/+] [left] OHCs). Insets [middle panel]: enlarged views of individual peroxisomes. In *Pjvk*[+/+] OHCs, peroxisomes are grouped just under the cuticular plate (CP) (arrowheads), with none detected in the perinuclear region (n = 33 sections, upper bar chart). In *Pjvk*[−/−] OHCs, some peroxisomes remain under the CP (arrowheads), but catalase-containing structures, misshapen peroxisomes (arrows), are detected in the perinuclear region (n = 24 sections, upper bar chart). Peroxisomes located under the CP are larger in *Pjvk*[−/−] OHCs (n = 92 peroxisomes) than in *Pjvk*[+/+] OHCs (n = 89 peroxisomes) (lower bar chart). N, cell nucleus. **p < 0.01, ***p < 0.001. Error bars represent the SEM. Scale bars are 5 μm in (A) and (B) and 0.5 μm in (E).

See also Figures S6 and S7.

Figure 7. Therapeutic Approaches in *Pjvk*^{−/−} Mice

(A–C) Effect of N-acetyl cysteine (NAC) on auditory function in *Pjvk*^{−/−} mice. (A) ABR thresholds in untreated versus NAC-treated P21 *Pjvk*^{−/−} mice. (B) ABR wave I amplitude for 10 kHz tone bursts in *Pjvk*^{+/+} and untreated *Pjvk*^{−/−} versus NAC-treated *Pjvk*^{−/−} mice at P21. (C) EEBR wave E IV amplitude before (dots) and after (crosses) controlled electrical stimulation of the cochlear nerve at 200 impulses/s for 1 min in *Pjvk*^{+/+}, untreated *Pjvk*^{−/−}, and NAC-treated *Pjvk*^{−/−} mice.

(D–I) Effect of AAV2/8-Pjvk-IRES-EGFP transferred into the cochlear hair cells on the auditory function of *Pjvk*^{−/−} mice. See also Figure S2B. (D) ABR thresholds at 10, 15, and 20 kHz in AAV2/8-Pjvk-IRES-EGFP-treated versus untreated *Pjvk*^{−/−} mice. (E and H) DPOAE threshold (E) and ABR wave I amplitude (H) at 10 kHz in

(legend continued on next page)

respectively; t test, p = 0.17), and no enlarged PMP70-stained structures were detected (data not shown). The absence of pejvakin thus resulted in defective sound-induced peroxisomal proliferation (both at 105 dB SPL and 97 dB SPL) and, even, in peroxisome degeneration (at 105 dB SPL) in IHCs. In $Pjvk^{-/-}$ mice injected with AAV2/8-Pjvk-IRES-EGFP on P3 and exposed to 105 dB SPL for 1 hr on P21, enlarged PMP70-labeled structures were no longer detected in transduced IHCs, and the number of peroxisomes increased by 35% (32.3 ± 2.1 and 43.7 ± 3.0 peroxisomes per IHC, n = 60 cells from unexposed and exposed transduced $Pjvk^{-/-}$ IHCs, respectively; t test, p = 0.002) (Figure 7J). We conclude that pejvakin re-expression fully protects $Pjvk^{-/-}$ IHCs from the degenerescence of peroxisomes and partially restores their impaired adaptive proliferation.

DISCUSSION

Noise-induced hearing loss (NIHL) is the second most common form of sensorineural hearing impairment after presbycusis in the United States (Dobie, 2008). Here, we describe a genetic form of NIHL, by showing that pejvakin deficiency in mice and DFNB59 patients leads to hypervulnerability to sound, due to a peroxisomal deficiency. To our knowledge, a peroxisomal cause of an isolated (non-syndromic) form of inherited deafness has not been reported yet. The peroxisome emerges as a key organelle in the redox homeostasis of the auditory system, for coping with the overproduction of ROS induced by high levels of acoustic energy.

Acoustic energy is the main determinant of NIHL. The $L_{EX,8\ hr}$ (for level of exposure over an 8-hr workshift) index has been defined such that an $L_{EX,8\ hr}$ of X dB delivers the same energy as a stable sound of X dB played over a period of 8 hr. Chronic occupational exposures to less than 85 dB (or 80 dB, depending on the country) are deemed safe. In $Pjvk^{-/-}$ mice, a single exposure to 63 dB $L_{EX,8\ hr}$ increased hearing thresholds by 30 dB, with full recovery occurring after about 2 weeks. By contrast, a ten times more energetic exposure to a $L_{EX,8\ hr}$ of 73 dB in wild-type mice of the same strain produces only an 18 dB shift in threshold, with a recovery time of 12 hr (Housley et al., 2013). This hypersensitivity of $Pjvk^{-/-}$ mice to noise suggests that the $L_{EX,8\ hr}$ of about 83 dB for a cage of ten pups is sufficient to account for permanent hearing loss in these $Pjvk^{-/-}$ pups, while some of those housed in small numbers in quiet rooms can display near-normal hearing thresholds (see Figure 1C). Likewise, the auditory function of DFNB59 patients was transiently affected by a 57 dB $L_{EX,8\ hr}$ exposure, routinely used in ABR tests.

NIHL involves the excessive production of ROS, overwhelming the antioxidant defense system and causing irreversible oxidative damage to DNA, proteins, and lipids within the cell (Henderson et al., 2006). Noise-induced oxidative stress results in the production of H_2O_2 and other ROS as by-products, thought to derive from the intense solicitation of mitochondrial activity, and several mouse mutants with mitochondrial defects are prone to NIHL (Ohlemiller et al., 1999; Brown et al., 2014). Our studies of pejvakin-deficient mouse mutants and rescue experiments targeting the hair cells and auditory neurons unambiguously show that IHCs, OHCs, primary auditory neurons, and neurons of the cochlear nucleus are hypervulnerable to sound in the absence of pejvakin, which is consistent with previous results showing that hair cells and neurons of the auditory system are targets of NIHL (Wang et al., 2002; Kujawa and Liberman, 2009; Imig and Durham, 2005). However, our study goes one step further by implicating a possible common mechanism: peroxisomal failure, the importance of which is demonstrated by the impairment of the redox homeostasis caused by pejvakin deficiency. It also reveals a major cause of the unusually high level of phenotypic variability observed in pejvakin-deficient mice and humans: the difference in sound exposure and the inability of the peroxisomes to cope with the resultant activity-dependent oxidative stress in the absence of pejvakin. Incidentally, this can account for the apparent paradox that mice carrying the R183W mutation in pejvakin displayed a much more severe neural pathway defect than the $Pjvk^{-/-}$ mice (Delmaghani et al., 2006). Due to the preservation of hair cell functions, the auditory neurons of R183W mutant mice should be strongly stimulated, whereas the early permanent damage to cochlear hair cells in $Pjvk^{-/-}$ mice acts as a protective "muffler" of the neuronal pathway.

In mammals, the number and metabolic functions of peroxisomes differ between cell types. However, all cell types are able to adapt rapidly to modifications in physiological conditions by changing the number, shape, size, and molecular content of peroxisomes, resulting in considerable functional plasticity of these organelles (Schrader et al., 2012; Smith and Aitchison, 2013). Our experiments on $Pjvk^{-/-}$ and $Pjvk^{+/+}$ mouse embryonic fibroblasts stressed with H_2O_2 showed that pejvakin is critically involved in the oxidative stress-induced proliferation of peroxisomes through growth and fission of pre-existing peroxisomes. The molecular machinery underlying this adaptive process is still poorly understood beyond the involvement of Pex11α (Li et al., 2002). Of note, the absence of pejvakin only affects the proliferation of peroxisomes from pre-existing peroxisomes, but not the constitutive biogenesis of this organelle. Accordingly, structural abnormalities of peroxisomes in $Pjvk^{-/-}$ mice became apparent only after hearing onset, in the context of the oxidative stress produced by noise exposure. By contrast, the *PEX* gene defects causing Zellweger syndrome spectrum (ZSS) disorders (Waterham and Ebberink, 2012) and rhizomelic chondrodysplasia punctata affect the constitutive biogenesis of peroxisomes. Hearing impairment in ZSS disorders involves a severe impairment of neuronal conduction and has been attributed to defects in the synthesis of two essential myelin sheath components—plasmalogens and docosahexaenoic acid—which is critically dependent on peroxisomes. Our results suggest that ZSS also

---

treated versus untreated contralateral ears. (G) ABR wave I latency in treated versus untreated contralateral ears. (F) Correlation between DPOAE thresholds and the proportion of EGFP-tagged (i.e., transduced) OHCs. Six untreated ears have no recordable DPOAE (threshold arbitrarily set at 80 dB SPL; red diamond). (I) Correlation between ABR wave I amplitude at 10 kHz, 105 dB SPL, and the percentage of transduced IHCs (EGFP tagged).
(J) Effect of AAV2/8-Pjvk-IRES-EGFP on the peroxisomes in $Pjvk^{-/-}$ IHCs. Upper and lower panels show and quantify (bar charts) the peroxisomes in untreated mice 48 hr after sound exposure (5–20 kHz, 105 dB SPL for 1 hr) (peroxisome abnormalities are indicated by arrowheads). Error bars represent the SD in (A–I) and the SEM in (J). ns, not significant; *p < 0.05, **p < 0.01, ***p < 0.001.

includes a defective redox balance in the hair cells and neurons of the auditory system.

In the context of noise exposure, the upregulation of *Pjvk* transcription in the cochlea and the subsequent peroxisome proliferation in the hair cells and auditory neurons of wild-type mice suggest that pejvakin-dependent peroxisome proliferation in the auditory system is part of the physiological response to high levels of acoustic energy that result in increased amounts of ROS. This and the marked oxidative stress detected in the $Pjvk^{-/-}$ cochlea imply that the proliferation of peroxisomes plays an antioxidant role, similar to that reported in other cell types (Santos et al., 2005; Diano et al., 2011). The rapid elevation of the hearing threshold in $Pjvk^{-/-}$ mice in response to low-energy sounds and the increase in interwave I-IV latency observed in DFNB59 patients within a few seconds are consistent with an activity-dependent H_2O_2 production that, due to impaired cellular redox homeostasis, results in concentrations of H_2O_2 high enough to impact on the activity of various target proteins including ion channels and transporters (Rice, 2011). The worsening of hearing sensitivity, 2 days later, in the mutant mice lacking pejvakin, exacerbated by putting back the mice in a noisy environment, fits the picture of the absence of sound-induced biogenesis of peroxisomes (with their degeneration occurring in a high acoustic energy environment). We thus conclude that the hypervulnerability of $Pjvk^{-/-}$ mice and DFNB59 patients to sound does not result simply from an exacerbation, by sound, of a pre-existing redox-balance defect, but is the consequence of impaired adaptive proliferation of peroxisomes in the absence of pejvakin. Both defective peroxisome proliferation in IHCs of $Pjvk^{-/-}$ mice in response to sound exposure and its partial recovery by pejvakin cDNA transfer support this conclusion. A full recovery of the adaptive peroxisome proliferation produced by sound exposure may require higher concentrations of pejvakin or the sound-induced modulation of *Pjvk* transcription (see Figure 6A), which was missing in our rescue experiments (pejvakin cDNA expression being driven by a constitutive promoter).

In patients with hearing impairment, the amplification of sound by hearing aids or direct electrical stimulation of the auditory nerve by a cochlear implant delivers a stimulus with an energy level similar to that shown here to worsen the hearing impairment of $Pjvk^{-/-}$ mice within 1 min of sound exposure. Therefore, in cases of peroxisomal deficiency, as in DFNB59, specific protection against redox homeostasis failure is essential. Patients with such conditions should avoid noisy environments and a beneficial effect of hearing devices should require an antioxidant protection. N-acetyl cysteine was the only antioxidant drug tested here to display some, albeit limited, efficacy. By contrast, AAV-mediated gene therapy could potentially provide full protection. Finally, deciphering the sound-stress-induced protective signaling pathway involving pejvakin might lead to the discovery of therapeutic agents for NIHL.

EXPERIMENTAL PROCEDURES

Audiological Studies in Mice

Auditory tests were performed in an anechoic room, on anesthetized animals whose core temperatures were maintained at 37°C (see the Supplemental Experimental Procedures).

Audiological Tests in Patients

Informed consent was obtained from all the subjects included in the study. Pure-tone audiometry was performed with air- and bone-transmitted tones. Hearing impairment was assessed objectively, by measuring ABRs and transient-evoked otoacoustic emissions (TEOAEs). The nonlinear TEOAE recording procedure was used (derived from the ILO88 system), making it possible to extract TEOAEs from linear reflection artifacts from the middle ear, and to evaluate background noise. TEOAE responses were analyzed in 1-kHz-wide bands centered on 1, 2, 3 and 4 kHz.

Generation of an Anti-pejvakin Monoclonal Antibody

The 3′ end of the coding sequence of the Pjvk cDNA (NCBI:NM_001080711.2) was inserted into a pGST-parallel-2 vector (derived from pGEX-4T-1; Amersham). The resultant construct, encoding the C-terminal region of pejvakin (residues 290–352; RefSeq:NP_001074180.1) fused to an N-terminal glutathione S-transferase tag, was introduced into *Escherichia coli* BL21-Gold (DE3)-competent cells (Stratagene). The pejvakin protein fragment was purified on a glutathione-Sepharose 4B column, then subjected to size-exclusion chromatography and used as the antigen for immunization. Antibodies were produced by immunizing $Pjvk^{-/-}$ mice. An immunoglobulin G monoclonal antibody (K_D of 6×10^{-8} M), Pjvk-G21, was selected by ELISA on immunogen-coated plates.

Statistical Analyses

Quantitative data are presented as mean ± SD, unless otherwise mentioned. Statistical analyses were performed using GraphPad. Data were analyzed by paired or unpaired Student's t tests and, for multiple comparisons, either by one-way or two-way ANOVA or by t tests with the Bonferroni correction. Statistical significance of the differences observed between groups is defined as $p < 0.05$.

ACCESSION NUMBER

The accession number for the transcriptomic data reported in this paper is deposited in GEO:GSE72722.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.023.

AUTHOR CONTRIBUTIONS

C.P. and P.A. designed the study. S. Delmaghani designed and performed most of the experiments. S. Dartevelle produced the Pjvk-G21 antibody. G.S. and S. Delmaghani performed the microarray experiments. I.P., J.D., and S. Delmaghani performed the cell biology experiments. N.T., M.T., M.L., S. Delmaghani, and S.S. performed the ultrastructural studies. S. Delmaghani, A.M., and A.E. performed the rescue experiments; M.B. and D.D. performed the ex vivo electrophysiology; and P.A. performed the in vivo electrophysiology analyses. T.Z., M.A., and E.S.V. studied the mitochondrial membrane potential. C.P., P.A., S. Delmaghani, J.D., and J.-P.H. wrote the manuscript.

ACKNOWLEDGMENTS

REFERENCES

Ashmore, J. (2008). Cochlear outer hair cell motility. Physiol. Rev. *88*, 173–210.

Avan, P., Büki, B., and Petit, C. (2013). Auditory distortions: origins and functions. Physiol. Rev. *93*, 1563–1619.

Borck, G., Rainshtein, L., Hellman-Aharony, S., Volk, A.E., Friedrich, K., Taub, E., Magal, N., Kanaan, M., Kubisch, C., Shohat, M., and Basel-Vanagaite, L. (2012). High frequency of autosomal-recessive DFNB59 hearing loss in an isolated Arab population in Israel. Clin. Genet. *82*, 271–276.

Brown, K.D., Maqsood, S., Huang, J.Y., Pan, Y., Harkcom, W., Li, W., Sauve, A., Verdin, E., and Jaffrey, S.R. (2014). Activation of SIRT3 by the NAD$^+$ precursor nicotinamide riboside protects from noise-induced hearing loss. Cell Metab. *20*, 1059–1068.

Collin, R.W., Kalay, E., Oostrik, J., Caylan, R., Wollnik, B., Arslan, S., den Hollander, A.I., Birinci, Y., Lichtner, P., Strom, T.M., et al. (2007). Involvement of DFNB59 mutations in autosomal recessive nonsyndromic hearing impairment. Hum. Mutat. *28*, 718–723.

Delmaghani, S., del Castillo, F.J., Michel, V., Leibovici, M., Aghaie, A., Ron, U., Van Laer, L., Ben-Tal, N., Van Camp, G., Weil, D., et al. (2006). Mutations in the gene encoding pejvakin, a newly identified protein of the afferent auditory pathway, cause DFNB59 auditory neuropathy. Nat. Genet. *38*, 770–778.

Diano, S., Liu, Z.W., Jeong, J.K., Dietrich, M.O., Ruan, H.B., Kim, E., Suyama, S., Kelly, K., Gyengesi, E., Arbiser, J.L., et al. (2011). Peroxisome proliferation-associated control of reactive oxygen species sets melanocortin tone and feeding in diet-induced obesity. Nat. Med. *17*, 1121–1127.

Dobie, R.A. (2008). The burdens of age-related and occupational noise-induced hearing loss in the United States. Ear Hear. *29*, 565–577.

Ebberink, M.S., Koster, J., Visser, G., Spronsen, Fv., Stolte-Dijkstra, I., Smit, G.P., Fock, J.M., Kemp, S., Wanders, R.J., and Waterham, H.R. (2012). A novel defect of peroxisome division due to a homozygous non-sense mutation in the PEX11β gene. J. Med. Genet. *49*, 307–313.

Ebermann, I., Walger, M., Scholl, H.P., Charbel Issa, P., Lüke, C., Nürnberg, G., Lang-Roth, R., Becker, C., Nürnberg, P., and Bolz, H.J. (2007). Truncating mutation of the DFNB59 gene causes cochlear hearing impairment and central vestibular dysfunction. Hum. Mutat. *28*, 571–577.

Henderson, D., Bielefeld, E.C., Harris, K.C., and Hu, B.H. (2006). The role of oxidative stress in noise-induced hearing loss. Ear Hear. *27*, 1–19.

Housley, G.D., Morton-Jones, R., Vlajkovic, S.M., Telang, R.S., Paramananthasivam, V., Tadros, S.F., Wong, A.C., Froud, K.E., Cederholm, J.M., Sivakumaran, Y., et al. (2013). ATP-gated ion channels mediate adaptation to elevated sound levels. Proc. Natl. Acad. Sci. USA *110*, 7494–7499.

Imig, T.J., and Durham, D. (2005). Effect of unilateral noise exposure on the tonotopic distribution of spontaneous activity in the cochlear nucleus and inferior colliculus in the cortically intact and decorticate rat. J. Comp. Neurol. *490*, 391–413.

Kujawa, S.G., and Liberman, M.C. (2009). Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. J. Neurosci. *29*, 14077–14085.

Li, X., Baumgart, E., Dong, G.X., Morrell, J.C., Jimenez-Sanchez, G., Valle, D., Smith, K.D., and Gould, S.J. (2002). PEX11alpha is required for peroxisome proliferation in response to 4-phenylbutyrate but is dispensable for peroxisome proliferator-activated receptor alpha-mediated peroxisome proliferation. Mol. Cell. Biol. *22*, 8226–8240.

Lopez-Huertas, E., Charlton, W.L., Johnson, B., Graham, I.A., and Baker, A. (2000). Stress induces peroxisome biogenesis genes. EMBO J. *19*, 6770–6777.

Mizuno, Y., Kurochkin, I.V., Herberth, M., Okazaki, Y., and Schönbach, C. (2008). Predicted mouse peroxisome-targeted proteins and their actual subcellular locations. BMC Bioinformatics *9* (*Suppl 12*), S16.

Møller, A.R., and Jannetta, P.J. (1983). Interpretation of brainstem auditory evoked potentials: results from intracranial recordings in humans. Scand. Audiol. *12*, 125–133.

Ohlemiller, K.K., Wright, J.S., and Dugan, L.L. (1999). Early elevation of cochlear reactive oxygen species following noise exposure. Audiol. Neurootol. *4*, 229–236.

Oliver, D., Taberner, A.M., Thurm, H., Sausbier, M., Arntz, C., Ruth, P., Fakler, B., and Liberman, M.C. (2006). The role of BKCa channels in electrical signal encoding in the mammalian auditory periphery. J. Neurosci. *26*, 6181–6189.

Rice, M.E. (2011). H_2O_2: a dynamic neuromodulator. Neuroscientist *17*, 389–406.

Robles, L., and Ruggero, M.A. (2001). Mechanics of the mammalian cochlea. Physiol. Rev. *81*, 1305–1352.

Santos, M.J., Quintanilla, R.A., Toro, A., Grandy, R., Dinamarca, M.C., Godoy, J.A., and Inestrosa, N.C. (2005). Peroxisomal proliferation protects from beta-amyloid neurodegeneration. J. Biol. Chem. *280*, 41057–41068.

Schrader, M., and Fahimi, H.D. (2006). Peroxisomes and oxidative stress. Biochim. Biophys. Acta *1763*, 1755–1766.

Schrader, M., Bonekamp, N.A., and Islinger, M. (2012). Fission and proliferation of peroxisomes. Biochim. Biophys. Acta *1822*, 1343–1357.

Schwander, M., Sczaniecka, A., Grillet, N., Bailey, J.S., Avenarius, M., Najmabadi, H., Steffy, B.M., Federe, G.C., Lagler, E.A., Banan, R., et al. (2007). A forward genetics screen in mice identifies recessive deafness traits and reveals that pejvakin is essential for outer hair cell function. J. Neurosci. *27*, 2163–2175.

Smith, J.J., and Aitchison, J.D. (2013). Peroxisomes take shape. Nat. Rev. Mol. Cell Biol. *14*, 803–817.

Starr, A., and Rance, G. (2015). Auditory neuropathy. Handb. Clin. Neurol. *129*, 495–508.

Tang, X.D., Garcia, M.L., Heinemann, S.H., and Hoshi, T. (2004). Reactive oxygen species impair Slo1 BK channel function by altering cysteine-mediated calcium sensing. Nat. Struct. Mol. Biol. *11*, 171–178.

Wang, Y., Hirose, K., and Liberman, M.C. (2002). Dynamics of noise-induced cellular injury and repair in the mouse cochlea. J. Assoc. Res. Otolaryngol. *3*, 248–268.

Waterham, H.R., and Ebberink, M.S. (2012). Genetics and molecular basis of human peroxisome biogenesis disorders. Biochim. Biophys. Acta *1822*, 1430–1441.

Relaxation of Loaded ESCRT-III Spiral Springs Drives Membrane Deformation

Graphical Abstract
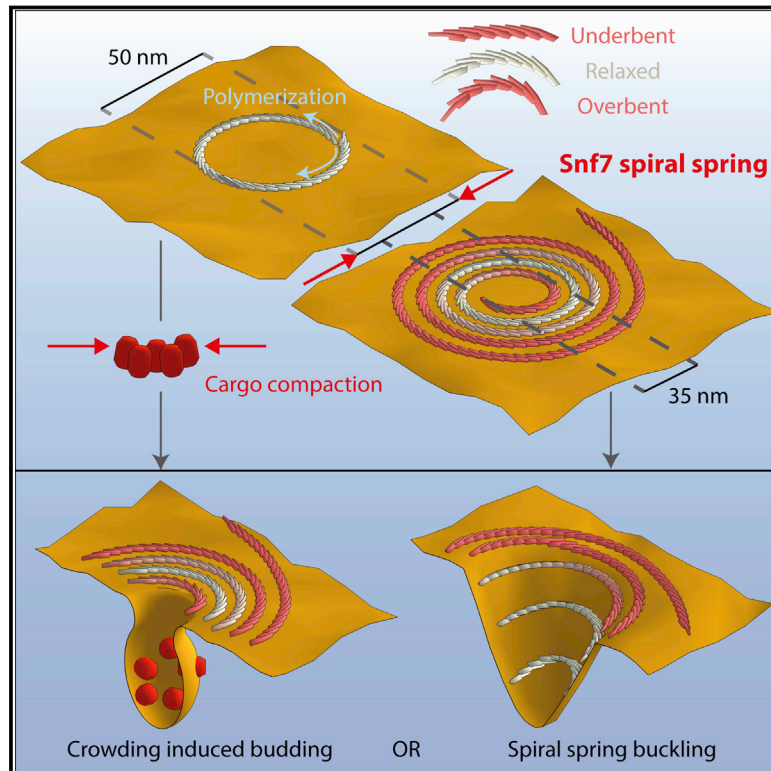
Authors

Nicolas Chiaruttini, Lorena Redondo-Morata, Adai Colom, Frédéric Humbert, Martin Lenz, Simon Scheuring, Aurélien Roux

Correspondence

simon.scheuring@inserm.fr (S.S.), aurelien.roux@unige.ch (A.R.)

In Brief

A component of the ESCRT-III membrane fission machinery self-organizes into spiral springs that trigger membrane deformation when released.

Highlights

- Snf7 forms highly flexible filaments that spontaneously curl

- Snf7 filaments forms spirals at the surface of lipid membranes

- Snf7 spirals are springs as they can deform under lateral compression

- Relaxation of compressed Snf7 spirals leads to membrane deformation

Relaxation of Loaded ESCRT-III Spiral Springs Drives Membrane Deformation

Nicolas Chiaruttini,[1,5] Lorena Redondo-Morata,[3,5] Adai Colom,[1,2,3] Frédéric Humbert,[1] Martin Lenz,[4,6] Simon Scheuring,[3,6,*] and Aurélien Roux[1,2,6,*]

[1]University of Geneva, Department of Biochemistry, quai Ernest Ansermet 30, 1211 Geneva 4, Switzerland
[2]Swiss National Centre for Competence in Research Programme Chemical Biology, 1211 Geneva, Switzerland
[3]U1006 INSERM, Aix-Marseille Université, Parc Scientifique et Technologique de Luminy, 163 avenue de Luminy, 13009 Marseille, France
[4]LPTMS,CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91405 Orsay, France
[5]Co-first author
[6]Co-senior author
*Correspondence: simon.scheuring@inserm.fr (S.S.), aurelien.roux@unige.ch (A.R.)

SUMMARY

ESCRT-III is required for lipid membrane remodeling in many cellular processes, from abscission to viral budding and multi-vesicular body biogenesis. However, how ESCRT-III polymerization generates membrane curvature remains debated. Here, we show that Snf7, the main component of ESCRT-III, polymerizes into spirals at the surface of lipid bilayers. When covering the entire membrane surface, these spirals stopped growing when densely packed: they had a polygonal shape, suggesting that lateral compression could deform them. We reasoned that Snf7 spirals could function as spiral springs. By measuring the polymerization energy and the rigidity of Snf7 filaments, we showed that they were deformed while growing in a confined area. Furthermore, we observed that the elastic expansion of compressed Snf7 spirals generated an area difference between the two sides of the membrane and thus curvature. This spring-like activity underlies the driving force by which ESCRT-III could mediate membrane deformation and fission.

INTRODUCTION

ESCRT-III (endosomal sorting complex required for transport) has been implicated in the formation of intraluminal vesicles (ILVs) during biogenesis of multi-vesicular bodies (MVBs) by genetic (Babst et al., 2002; Coonrod and Stevens, 2010) and biochemical assays (Adell et al., 2014; Henne et al., 2012; Saksena et al., 2009; Wollert and Hurley, 2010; Wollert et al., 2009). ESCRT-III budding occurs in an opposite direction than in endocytosis: the limiting membrane is pushed outward from the cytoplasm instead of curving inward. ESCRT-III has been proposed to play a role in membrane deformation (Hanson et al., 2008) and fission of ILVs (Adell et al., 2014). Consistent with this, ESCRT-III is also required for geometrically similar fission reactions

such as viral budding (von Schwedler et al., 2003) and abscission during cytokinesis (Carlton et al., 2008; Elia et al., 2011; Guizetti et al., 2011). ESCRT-III nucleation is promoted by ESCRT-II and its disassembly by the ATPase Vps4 (Lata et al., 2008).

It is unclear how ESCRT-III deforms lipid membranes. Because of their polymerization abilities, ESCRT-III proteins (Vps20, Snf7, Vps2, Vps24) have been proposed to generate membrane curvature by scaffolding (Cashikar et al., 2014; Fabrikant et al., 2009; Hanson et al., 2008; Lata et al., 2008). In this mode, polymers coating the membrane usually adopt a single specific shape, or, at least, a set of geometrically similar shapes. ESCRT-III filaments adopt instead a wide variety of shapes in vivo and in vitro: concentric circles, rings, spirals, helices, or linear filaments have been observed (Hanson et al., 2008; Henne et al., 2012; Pires et al., 2009). Furthermore, no unique shape for the assembly of ESCRT-III proteins arises from the molecular structure of ESCRT-III proteins (McCullough et al., 2013). Instead, curvature could be generated by other mechanisms: for example, it has been proposed that the amphipathic insertion of the N-terminal part of Snf7 could participate in the generation of membrane curvature (Buchkovich et al., 2013). We were thus interested in studying how ESCRT-III polymerization could drive membrane curvature.

RESULTS

Growth of Snf7 Patches on Supported Bilayers

To study the polymerization of ESCRT-III, we reconstituted ESCRT-III polymerization by adding purified yeast Snf7 onto supported lipid membranes. Supported membranes were obtained by bursting giant unilamellar vesicles (GUVs) composed of 40% di-oleoyl-phosphatidylserine (DOPS) and 60% of di-oleoyl-phosphatidylcholine (DOPC) on cleaned glass coverslips (Figure S1A, Movie S1). These coverslips were built into a flow chamber, allowing sequential addition and exchange of solutions.

First, Snf7 labeled with Alexa488 (Snf7-Alexa488) was flushed into the chamber, and its association to the membrane was imaged by time-lapse spinning-disk confocal microscopy (SDC). At 400 nM, Snf7 formed patches evenly distributed on
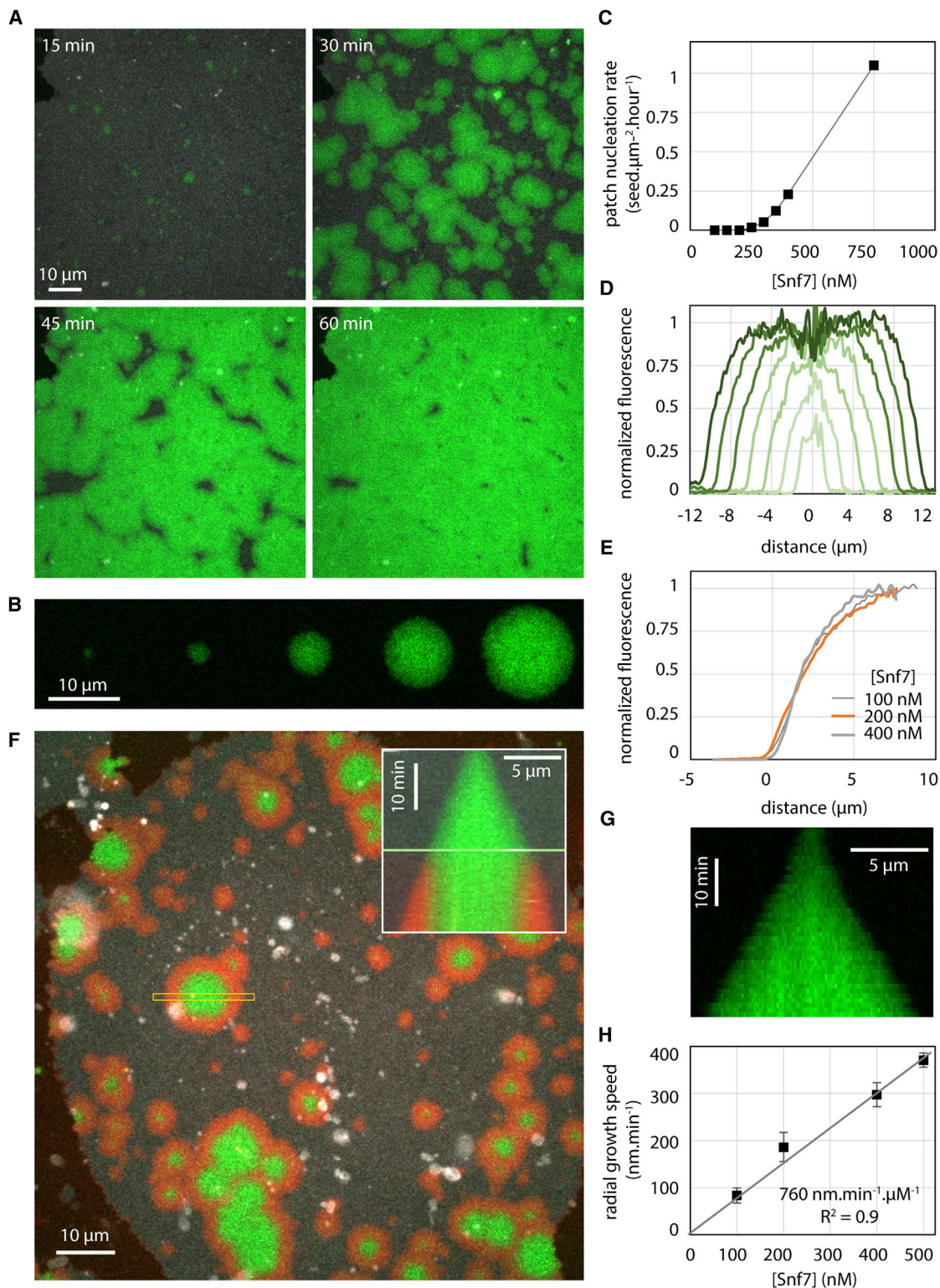
Figure 1. Nucleation and Growth of Snf7 Patches on Supported Membranes

Lipid composition is DOPC 60% / DOPS 40%+ Rhodamine PE 0.1%.

(A) Time-lapse images of Snf7-Alexa488 patches growth (green) at [Snf7] = 400 nM on supported membrane (gray).

(B) Time-lapse images (every 10 min) of a single Snf7-Alexa488 patch (green) growing at [Snf7] = 200 nM.

the membrane surface (Figure 1A). These circular patches (Figure 1B) grew over the course of an hour until becoming confluent, eventually covering the membrane completely.

The nucleation rate of the patches depended on Snf7 bulk concentration (Figure 1C). Patch formation was not observed below 200 nM. Above 1 μM, patch formation and growth was so fast that individual patches could hardly be discriminated. Between these limiting concentrations, we were able to follow micron-sized patches individually over several tens of minutes (Movie S2). Once formed, patches disassembled with a half time of approximately 15 hr upon Snf7 washout (Figure S1B; Movie S3).

We termed "patch nucleation" this nucleation of patches in the absence of a previous Snf7 structure. The patch nucleation rate was very low, less than 1 seed.μm^{-2}.$hour^{-1}$, and depended on the amount of negatively charged lipids in the membrane (Figure S1C), revealing the critical role of these lipids in promoting Snf7 polymerization. We did not observe Snf7 assemblies in the absence of membranes.

The periphery of the patches showed dimmer Snf7 fluorescence than the center: fluorescence decayed radially at the rim over the outer 4 μm. This gradient of fluorescence was the same as the patch grew in radius (Figure 1D) and independent on bulk concentration of Snf7 (Figure 1E). The central part of the patches had the same intensity, constant over time. We postulated that the patch could be made of two parts: a central part where Snf7 entirely covers the membrane and cannot further assemble, and a rim, representing a growing front. To study whether Snf7 was assembled only in the front region, we generated patches with a solution of Snf7-Alexa488, which we then replaced with Snf7-Atto647N (Figure 1F and Movie S4). As postulated, Snf7-Atto647N fluorescence appeared only at the border of the growing patches. These observations confirmed that Snf7 patches were growing by a traveling circular front.

The front propagated at constant speed (Figure 1G). The speed was linear with Snf7 concentration (Figure 1H). As a result, and because the fluorescence gradient was independent from Snf7 concentration, fluorescence intensity curves with time at a given point (Figure S1E) could be merged into a single one by rescaling time with Snf7 concentration (Figure S1F). The amount of negatively charged lipids also affected the front speed (Figure S1D). In summary, the growth of the Snf7 patches reflected a nucleation/growth process, with a nucleation rate of less than 1 seed.μm^{-2}.$hour^{-1}$ and a radial growth speed of 760 nm.min^{-1}.μM^{-1}.

Since Snf7 filaments can curl into circles or spirals (Hanson et al., 2008; Henne et al., 2012; Shen et al., 2014), we reasoned that one patch could be made of a single spiral filament growing from its tips at constant rate. In this case, however, the radial growth speed of the patch should slow down as the square root of time, and a dimmer fluorescence at the periphery of the patches would not be expected. To resolve this apparent contradiction, we studied the molecular structure of the Snf7 patches with atomic force microscopy (AFM) and electron microscopy (EM).

Snf7 Patches Are Made of Spiraling Filaments with Lateral Interactions

We first acquired images of Snf7 patches by AFM. GUVs composed of 60% DOPC and 40% DOPS were burst on a mica support (Figure S2A). After a four hours incubation of Snf7 at 1 μM, AFM images revealed that the micron-sized Snf7 patches consisted of packed arrays of Snf7 circular assemblies (Figure 2A; Figure S2B). Each assembly was formed by concentric circle-like structures. However, in these packed conditions, rather than being perfectly circular, each assembly was deformed into polygons with six neighbors on average (Figure 2B). The average external radius was 123 ± 35 nm (in the following, values are mean ± SD unless otherwise noted; n = 295) (Figure 2C), and the innermost circle had an average radius of 18 ± 3 nm (n = 120) (Figure 2D). The average distance between successive circles was $b = 17 \pm 3$ nm (n = 80) (Figure 2E).

To study the structure of these Snf7 assemblies, we performed negative stain electron microscopy (EM) of large unilamellar vesicles (LUVs) coated with Snf7 upon incubation for 15 min in a 1 μM Snf7 solution. LUVs coated with circular structures were observed, consistent with AFM images (Figure 2F). The fine structure of the filaments remained difficult to see because the two hemispheres of the LUV are projected onto the same EM image. However, in many cases, LUVs that had adhered on the grid were flushed during staining, leaving Snf7 assemblies attached to the grid surface. Two kinds of structures were then observed: small rings (27 ± 4 nm average radius; n = 61; Figure 2G) and large circular assemblies (R = 110 ± 40 nm; n = 46; Figure 2H).

Small rings are composed of filaments with two different thicknesses (Figure 2G and Figure S2C). The thinner ones appeared single stranded with an approximate thickness of 4.5 ± 0.3 nm (n = 10) in agreement with previous data (Pires et al., 2009; Shen et al., 2014). The thicker ones were double-stranded, with approximately twice the thickness (10.7 ± 0.7 nm; n = 7). The larger circular assemblies had an average radius of 110 ± 40 nm (n = 46). Following the path of the innermost filament in these assemblies revealed that they were made of a single filament, self-organized into a spiral (Figure 2H; Figure S2D). Within the spiral, we observed that the filament could associate laterally with itself, forming double-stranded filaments. These images showed that the spiraling nature of Snf7 filaments previously observed in solution (Shen et al., 2014) is conserved at the

(C) Patch nucleation rate as a function of [Snf7].

(D) Successive (from bright to dark green, every 10 min) Snf7 patch fluorescence profiles (circularly averaged) at [Snf7] = 200 nM.

(E) Snf7 patch edge fluorescence profile (average of 3 patches) as a function of [Snf7] (data for [Snf7] < 200 nM were obtained by first nucleating the patches at 350 nM for 5 min, and then [Snf7] was reduced to the desired value).

(F) Exchange of bulk Snf7-Alexa488 (green) with Snf7-Atto647N (red) at 200 nM. Inset: kymograph of the region selected (yellow box). The green line is the switch between green and red Snf7.

(G) Equatorial kymograph of the patch shown in B.

(H) Patch radial growth speed as a function of [Snf7]. The slope of the linear fit (gray line) is 760 nm.min^{-1}.μM^{-1}.
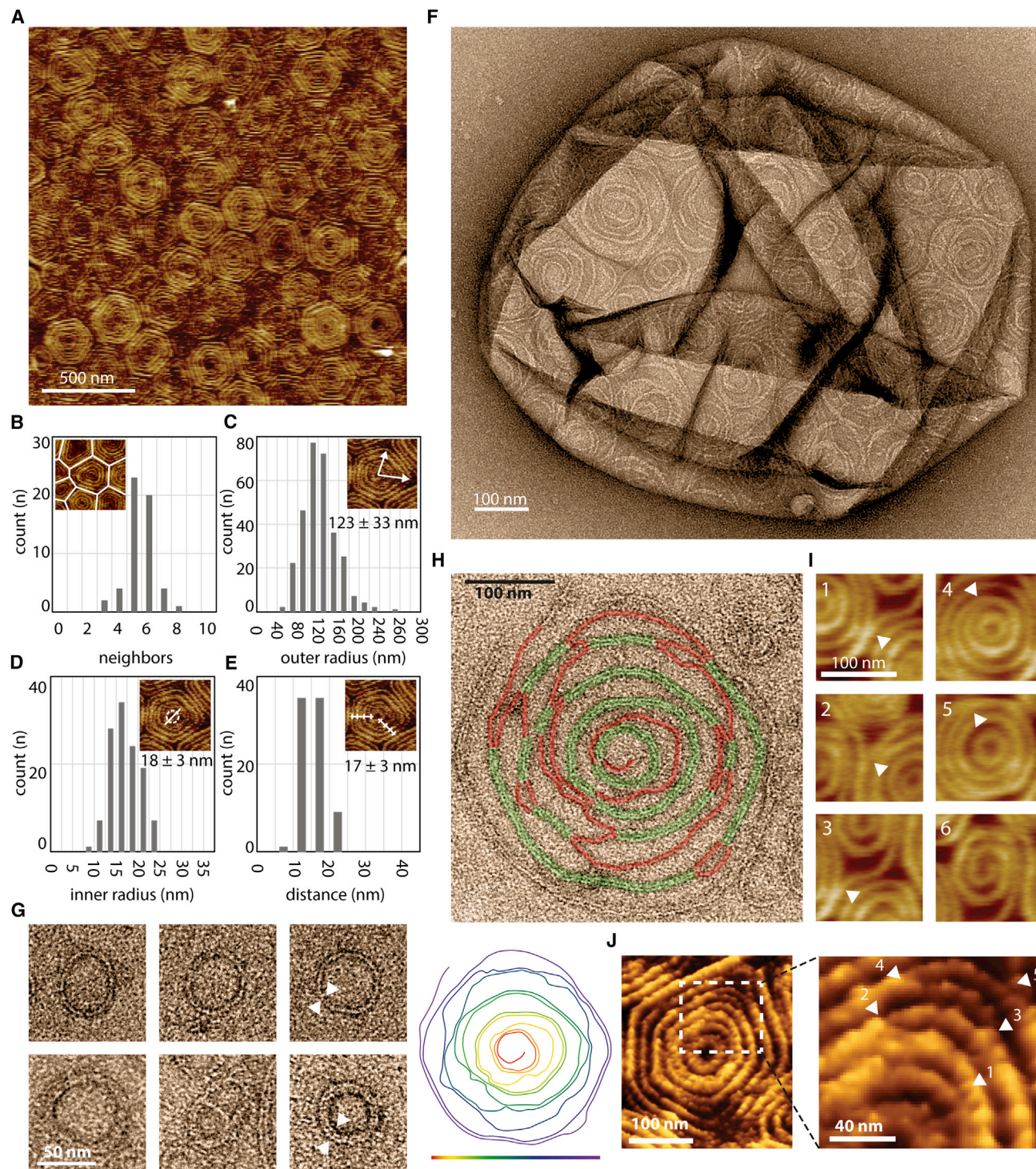
See also Figure S1.

Figure 2. Patches Are Made of Packed Snf7 Spirals

(A) AFM topographic image of the center of a Snf7 patch.

(B) Histogram of the number of neighbors per assembly.

(C) Histogram of the average outer radii of Snf7 assemblies. The average radius is 123 nm ± 33 nm.

(D) Histogram of the innermost circle radii. The average radius is 18 nm ± 3 nm.

(E) Histogram of the inter-circle distance. The average distance is b = 17 nm ± 3 nm.

(F) TEM image of a negatively stained, Snf7-coated LUV.

membrane. It also suggests that previously reported circular assemblies could also be spirals: in vitro, a Snf7 mutant protein that polymerizes spontaneously onto membranes (Henne et al., 2012) or, in fibroblasts, the circular structures found upon over-expression of CHMP4A and CHMP4B (Hanson et al., 2008). The formation of rings suggests that Snf7 filaments grow with a preferred curvature and the formation of spiral suggests that Snf7 filaments can depart from this preferred curvature with some flexibility.

The AFM images also revealed that filaments could split within the disk assemblies (Figure 2I): thick filaments that appeared as concentric circles are actually interconnected by thinner filaments (Figure 2J, arrows). This is consistent with the spiral structure seen by EM: double-stranded parts are combined with single-stranded connections (Figure 2H). The thinner filaments had an average thickness of 4.9 nm ± 1.5 nm (n = 25), probably corresponding to single strands, whereas thicker ones had a thickness of 10.6 ± 1.2 nm (n = 25), consistent with double-strands. The AFM and EM analysis suggested a structure of the Snf7 assemblies, where a single spiral filament interacted laterally with itself to form double-stranded filaments. Occasionally, spirals were directly observed by AFM (Figure 2I, subpanel 6).

These observations prompted the question of how Snf7 patches were formed. Patch nucleation could, for instance, start from a single closed ring, like those seen by EM. It is conceivable that such rings could be prone to break open, thus freeing filament tips that could further grow into a spiral. How would then this initial spiral transform into a patch? A possible scenario consists of a two-step growth mechanism (Figure 3A). First, new spirals are nucleated in the vicinity of existing spirals (termed below spiral nucleation). Rupture of filaments would separate the newly formed spirals from the initial spiral. Second, these spirals would grow independently through the addition of monomers at their filament tips. This scenario accounts for the observed growth dynamics of Snf7 patches: the constant speed of the radial growth of the patches implies that the density of growing filament tips at their rim stays constant. The formation of new spirals generates new tips, maintaining a constant density of growing tips.

To explore quantitatively the implications of such scenario, we developed a mathematical description of the dynamics of surface coverage by growing Snf7 spirals (Figure 3B and Figure S3). In the model, Snf7 spirals are represented by hard disks deposited on a surface representing a small (micron-sized) piece of membrane. As initial conditions, a few disks with radius r_0 are present, corresponding to events of initial patch nucleation. Patch nucleation is then neglected (set to zero) during the rest of the dynamics. New disks are thus generated only by spiral nucleation.

Disks growth corresponds to an area gain w per unit time (in $nm^2.s^{-1}$), as expected if the Snf7 filaments are elongating

from their tips at a constant speed w/b (in $nm.s^{-1}$, where $b = 17$ nm is the distance between Snf7 filaments in a single spiral). As the Snf7 spiral grows, its perimeter increases, offering an increasing number of potential spiral nucleation sites. We model this by stochastic nucleation of new disks with radius r_0 at a rate $\lambda \mathscr{P}$, where $\mathscr{P}$ denotes the total perimeter of the existing disks, and λ a constant spiral nucleation rate (expressed in number of nucleation events per second per micrometer). In the model, nucleation is prevented if the new disk location is already occupied by an existing disk. In addition, both nucleation and polymerization stop when the surface is completely covered with disks (Figure 3C). Solving the model in a mean-field approximation using the value obtained experimentally for r_0 gives a final distribution of disk sizes with one unknown parameter (w/λ) (Supplemental Information, Supplemental Mathematical Modeling part 1). Fitting $(w/\lambda) = 9.8 \pm 1.5 \times 10^{-3}$ μm^3, we find the distribution in excellent agreement with the experimental size distribution (Figure 3D).

We then tested experimentally whether three key features of our theoretical model are indeed fulfilled during the generation of Snf7 assemblies. First, both the existence of initial single rings and the growth into spirals imply that Snf7 filaments have a preferred high curvature. Since the average radius of rings is in the range 25–30 nm, this might correspond to the preferred curvature. Second, the proposal of secondary nucleation of spirals implies that new spirals can form from existing ones. Third, polymerization arrest should be correlated with contacts between neighboring disks. Finally, we also sought to determine independently the parameters w and λ, whose ratio (w/λ) had been estimated from the fit of the size distribution of spirals (Figure 3D). To address all this, we studied the molecular dynamics of spiral growth by total internal reflection fluorescence microscopy (TIRFM) and by dynamic high-speed AFM (HS-AFM) imaging (Casuso et al., 2012).

Snf7 Filament Dynamics Uncovers the Intrinsic Filament Curvature and the Mode of Nucleation

To characterize the initial events leading to patch formation, we first studied the early steps of Snf7 patch nucleation by TIRFM, which allowed us to quantify the approximate number of Snf7 molecules in diffraction limited spots from their fluorescence intensity (see Experimental Procedures for quantification). Patch nucleation started with the appearance of a fluorescent diffraction limited spot (nucleus, Figure 4A) containing 50 ± 20 monomers (n = 9). At a Snf7 concentration of 300 nM, the intensity of the nuclei remained constant for several minutes until these nuclei started to grow (Figure 4A). Under these conditions, the number of nuclei is very low (Figure 1C). To increase the number of nuclei and to obtain more robust statistics, we nucleated Snf7 assemblies by adding 1 μM ESCRT-II and 1 μM Vps20 to a 75 nM Snf7 solution. Under these conditions, many Snf7 nuclei

(G) TEM images of Snf7 rings, single (upper row) and double (lower row) stranded.

(H) Top: TEM image of a single Snf7 spiral. The Snf7 filament is underlined in green (resp. red) when double stranded (resp. single stranded). Bottom: color code of the filament path from the most inner turn (red) to the most outer turn (purple). See also Figure S2D.

(I) AFM images of connections between filaments: 1 to 3, split filaments connecting two spirals – 4 and 5, filament split within a spiral – 6, a spiral filament.

(J) High resolution AFM topographic image of Snf7 filament splitting and branching within a single Snf7 spiral.
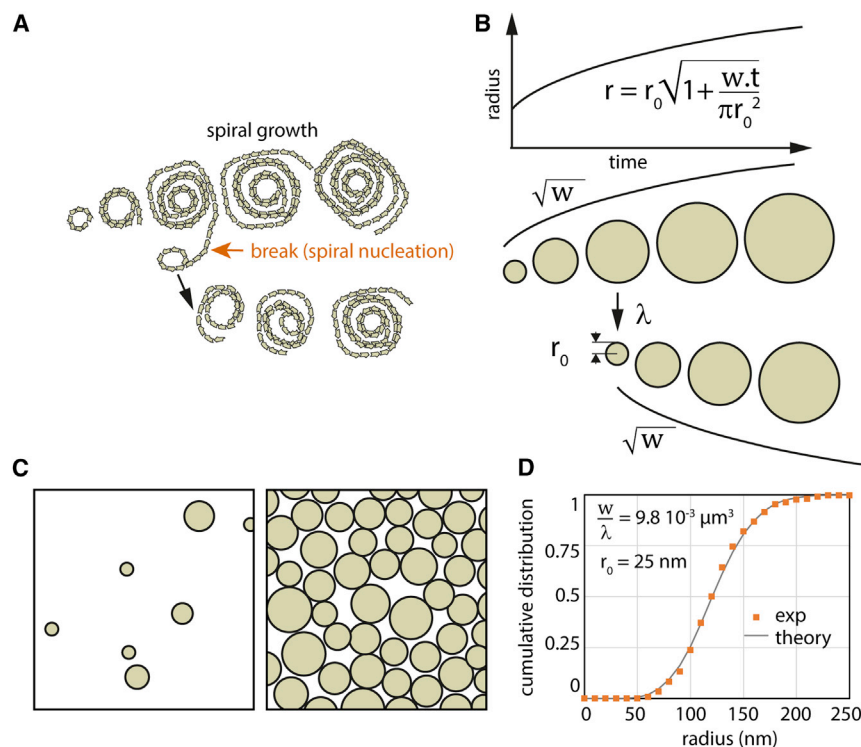
See also Figure S2.

Figure 3. Modeling of Snf7 Patch Growth

(A) A putative scenario for the nucleation and growth of Snf7 spirals into a patch: new spirals are formed from filaments protruding from pre-existing spirals. The new spirals separate from the mother spiral by filament break.

(B) Schematic of the theoretical model for Snf7 patch growth. Snf7 spirals are represented by disks. Disks are created with an initial radius r_0. Their area grows with a constant rate (w), which leads to a radius growing as the square root of time (upper graph and black curves). New spirals are nucleated over time proportionally to the spiral nucleation rate λ and to the total perimeter of existing disks.

(C) Pictorial representations of a small membrane area being covered with Snf7 disks at the beginning (left) and at the end (right) of the growth process.

(D) Cumulative distribution of spiral sizes (dots, calculated from Figure 2C) fitted with our theoretical model (line), imposing $r_0 = 25$ nm. The single fit parameter (w/λ) is equal to 9.8×10^{-3} μm^3.

appeared on the membrane surface (Figure 4B and Figure S4A) and remained stable for several tens of minutes, consistent with our observations with Snf7 alone (see Figure 4A).

The Snf7 nuclei have an average number of molecules of 60 ± 46 (mean $\pm$ SD; n = 1856, Figure 4C). If arranged in a closed ring, it would generate a circle of about 30 nm radius, considering a distance of 3.2 nm between Snf7 monomers in the filaments (Shen et al., 2014) (radius = perimeter / (2π) = $(60 \times 3.2)/(2 \times 3.14) \sim 32$ nm). This radius corresponds to those observed in rings seen by EM (see Figure 2G), implying that the arrested nuclei observed by TIRF could be closed rings. This supports that Snf7 patch nucleation starts by the appearance of a single, highly curved Snf7 ring that would break to form a spiral. To further confirm this hypothesis, we further showed through photobleaching experiments that breakage of the nuclei induces patch formation (Figures S4B and S4C).

The formation of highly curved nucleation rings suggests that Snf7 filaments have a preferred radius of curvature in the 25–30 nm range. Indeed, we measured radii of a 27 nm radius by EM and estimated a 32 nm radius by TIRF, in line with other studies (21 nm (Shen et al., 2014) and 32 nm (Henne et al., 2012)). In spirals, this preference should not be satisfied, as outer turns are under-curved (123 nm radius, Figure 2C) and inner turns are over-curved (18 nm radius, Figure 2D). These forced suboptimal radii of curvature may induce significant mechanical stresses in the Snf7 filaments, which might in turn underlie Snf7 biological function in membrane deformation.

To explore the existence of such internal stresses, we used the HS-AFM tip as a nanodissector (Scheuring et al., 2003), briefly applying strong forces to partially break a densely packed array

of Snf7 spirals. After breakage, 8 large disks were transformed into 29 small circles (Figure 4D), with a radius of 17 ± 5 nm (see size distribution Figure S4D). This is consistent with a scenario where Snf7 filaments were excised by rupture from the outer circles and curled back to a radius closer to their preferred radius. The nanodissector experiment indicated that even if polymerized at low radius of curvature, Snf7 filaments kept their ability to curl into smaller rings.

The radius of these broken filaments (17 nm) is smaller than the one of the initial ring ($\sim$27 nm), but very close to the size of the inner turns in large spirals (18 nm; see Figure 2D). This could be consistent with a preferred radius of curvature of 17 nm. Alternatively, the nanodissector-induced rings experienced the lateral pressure of the neighbors, forcing them to a smaller radius. To discriminate these two hypotheses, we studied the nucleation of new spirals from pre-existing Snf7 assemblies at molecular resolution by HS-AFM. We focused on areas where Snf7 spirals were already packed, but free membrane was still available (Figures 4E and F; Movies S5 and S6). Our image sequences showed that newly formed spirals were mainly initiated from filaments protruding from pre-existing spirals (Figure 4E), revealing the mechanism by which spiral nucleation occurs. The outer radius of these spirals grew with time while forming new turns (Figures 4F and 4G). While growing, bundled filaments can transiently separate and interact laterally with the neighboring bundles (Figure 4H).

Strikingly, when increasing from 2 to 3 concentric circles, the radius of the innermost circle was reduced from 22 nm to 14 nm (Figure 4G). These observations strongly support our hypothesis that the preferred radius of curvature of Snf7 oligomers is about 25 nm, but that lateral pressure can induce higher curvatures.

The radial growth of free spirals was initially rapid, but then slowed down (Figure 4I), roughly following a dependence on

the square root of time as expected from our model. From these data, we measured an area growth rate w of 80 ± 36 nm^2.s^{-1} (n = 5; see Experimental Procedures for quantification) at 1 μM. We estimated from w a growth rate of approximately 3 subunits.s^{-1}.μM^{-1} (Supplemental Information, Supplemental Mathematical Modeling). The estimated filament growth rate is in the range of other filament rates: actin and tubulin are in the range of 5–10 subunits.s^{-1}.μM^{-1}. Using our previous estimate $(w/\lambda) = 9.8 \times 10^{-3}$ μm^3 and this experimental value for w, we obtained the spiral nucleation rate as $\lambda = 8.2 \times 10^{-3}$ spiral.μm^{-1}.s^{-1} at [Snf7] = 1 μM. This secondary spiral nucleation rate is 500 times larger than the initial patch nucleation rate (Supplemental Information, Supplemental Mathematical Modeling). It validated our assumption that the patch nucleation was negligible in our theoretical model.

In summary, these observations indicated (1) a preferred high curvature of the Snf7 filaments and (2) the mechanism of spiral nucleation from existing spirals. We then set up to study the third feature of our model, whether lateral contacts between spirals can inhibit their growth.

Polymerization of Snf7 Filaments Induces Compression of the Spirals

AFM images of packed arrays of spirals showed that filaments at the contact zone between spirals were flattened, resulting in spirals acquiring a polygonal shape that was more pronounced for longer incubation times (Figure 5A). Moreover, the central area of some of these polygons was pushed toward the substrate, as seen in the height profile of AFM images (Figure 5B). Also, the centers of these spirals were often found to be stiffer as seen in AFM mechanical maps (Figure 5B). We reasoned that this deformation reflected lateral compression of the spirals as the membrane became covered with Snf7. To study the correlation between polymerization rate and lateral compression of the Snf7 assemblies, we ought to measure them simultaneously.

We started by reconstituting Snf7 polymerization on GUVs adhered to a glass surface, to avoid any displacement during time-lapse imaging (Figure S5A) and followed the increase of Snf7 fluorescence on the vesicle with time. In supported bilayers, which have smaller lipid mobility than free bilayers, diffusion of Snf7 assemblies is very limited. In contrast, Snf7 assemblies had a higher diffusion on GUVs explaining that fluorescence showed a homogeneous distribution (Figure 5C).

The dynamics of saturation of Snf7 polymerization over the entire GUV (Figure 5C, bottom) was similar to the dynamics of coverage at a single point in the supported bilayer experiments (Figure S5B): after an approximately exponential increase of the fluorescence, the dynamics of coverage saturated through a progressive slow down phase. If polymerization rate is independent of lateral compression, as it is in our theoretical model, an abrupt arrest of growth is expected (Figure S5C). Conversely, a polymerization rate dependent on lateral compression would cause a progressive slow down until reaching saturation.

As an indication of lateral compression, we noticed that coated GUVs underwent dramatic morphological changes upon several hours of Snf7 polymerization: GUVs were not spherical anymore and instead showed extreme irregular shapes, similar to rigid punched table tennis balls (Figure 5D).

Upon aspiration into a micropipette, they deformed plastically (Figure 5E and Figure S5D), showing that the Snf7 solidifies the membrane. These observations are consistent with a scenario where a rigid Snf7 coat generates pressure on the GUVs, stretching their membrane.

We reasoned that the accumulation of lateral compression in the Snf7 coat would stretch the underlying membrane, increasing its tension (Figure 5F). In order to follow the dynamics of accumulation of lateral compression within the Snf7 layer, we directly measured membrane tension generation during Snf7 polymerization. Using optical tweezers, we pulled a thin tether from a GUV held in an aspiration pipette, that allows to monitor membrane tension through the measurement of the force F exerted on the optical tweezers' bead (Cuvelier et al., 2005) (Figure 5G). We then flowed a 500 nM solution of Snf7 using an injection pipette, which triggered protein assembly onto the membrane (Figure 5H). Monitoring Snf7 fluorescence, we found that its membrane binding dynamics was identical to that measured previously, reaching the saturation value after 10–20 min (Figure 5I, top). Concomitant with this saturation, we observed an increase in the force exerted by the tube, indicating an increase of membrane tension (Figure 5I, bottom). These data show that Snf7 polymerization occurs at a slower rate as compression increases within the Snf7 layer, suggesting a coupling between polymerization and compression. The dependence of the polymerization energy μ (polymerization energy per unit surface) on the changes in tube force is captured by the expression (Supplemental Information, Supplemental Mathematical Modeling part 4.3):

$$\mu = \frac{F_f^2 - F_i^2}{8\pi^2 \kappa} \qquad (1)$$

where $F_i = 8$ pN and $F_f = 35$ pN are the tube forces before and after Snf7 polymerization and $\kappa = 4.8 \times 10^{-20}$ J, is the membrane bending rigidity (Figure S5E). These values yield $\mu = 3.1 \times 10^{-4}$ J.m^{-2}, which is a fairly high polymerization energy, twice that of clathrin (Saleem et al., 2015), which by itself is able to cause membrane deformation during endocytosis. This indicates that the polymerization force of Snf7 can plausibly cause membrane deformation.

These data are therefore compatible with a scenario where Snf7 spirals act as 2D springs that load themselves through polymerization. Constrained by their neighbors, Snf7 spirals would deform significantly during polymerization, from a disk-like to a polygonal shape, until deformation and polymerization force are balanced. To test quantitatively whether the polymerization force is sufficient to deform Snf7 spirals, we established a theoretical elastic model of spiral compression.

In the model, the Snf7 layer is approximated to a hexagonal lattice of individual spirals. Each spiral is composed of a collection of concentric filaments spaced by a distance $b = 17$ nm up to a typical radius $R = 130$ nm similar to the experimental data shown in Figure 2C. The filaments are circular in the absence of lateral compression, but may deform into hexagons with rounded vertices, to accommodate tighter packing, as illustrated in Figures 5A and 5F. The amount of energy required to deform circular filaments into more polygonal shapes depend on the
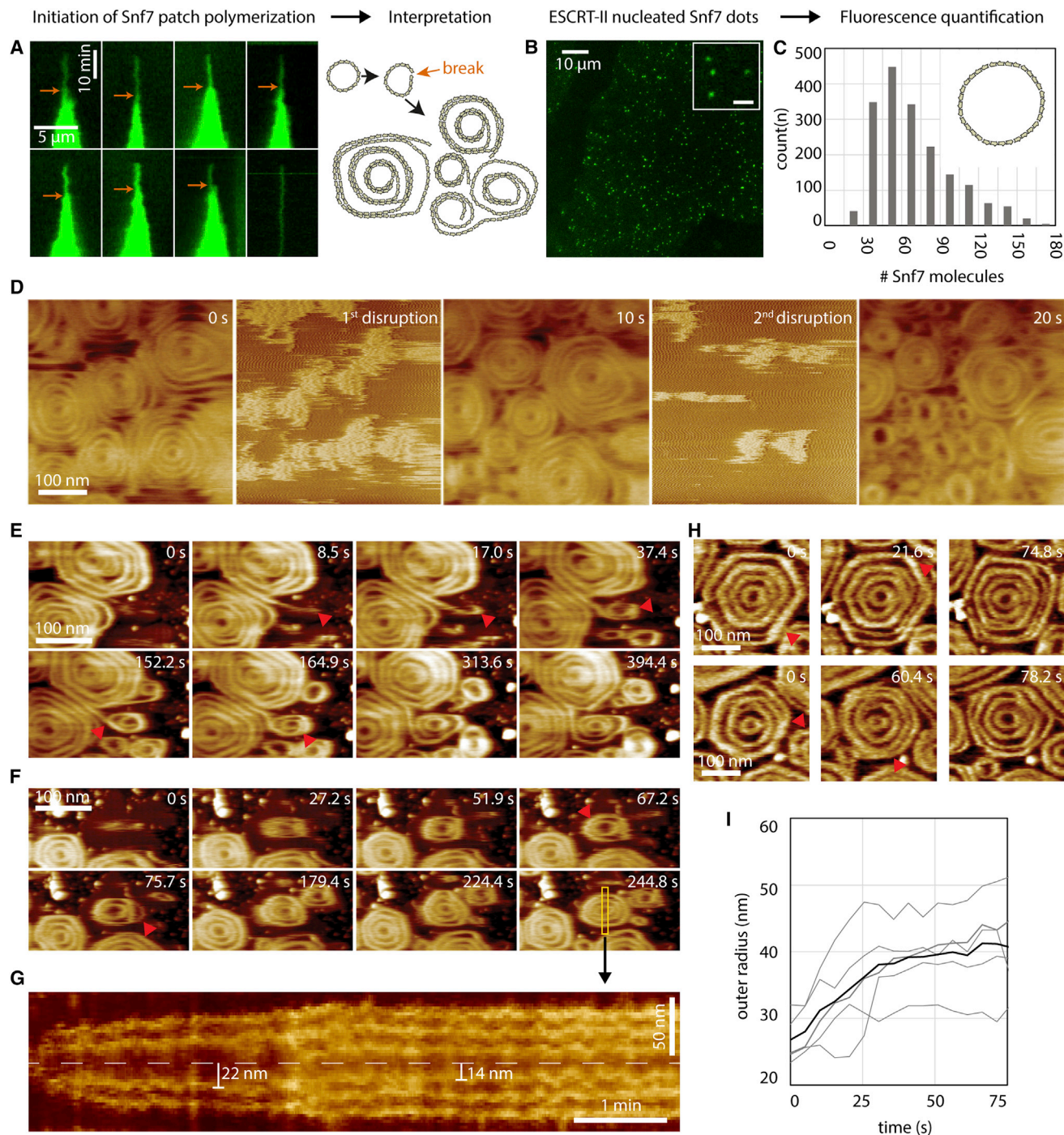
Figure 4. Nucleation and Growth of Snf7 Spirals on Supported Membranes

(A) Left: TIRF microscopy kymographs of the nucleation of single Snf7 patches (green) at [Snf7] = 300 nM. Arrows indicate single ring to multiple spirals transition as postulated from the interpretation of these observations (right).

(B) TIRF microscopy image of Snf7-Alexa488 dots (green) nucleated by ESCRT-II, [Snf7] = 75 nM, [Vps20] = 1 µM, [ESCRT-II] = 1 µM. Inset: zoom on 4 diffraction-limited spots (scale bar, 2 µm).

(C) Histogram of the estimated number of Snf7 molecules within the dots nucleated by ESCRT-II (n = 1856).

(D) HS-AFM nanodissection experiment (see text) of Snf7 spirals. 2 cycles of high AFM force were applied, between 0 s and 10 s, and between 10 s and 20 s.

(E) HS-AFM time-lapse sequence showing the apparition of a new Snf7 spiral from pre-existing ones. Arrowheads show: filament protruding from a spiral (t = 8.5 s), filament curling from its tip (t = 17.0 s), and forming a small spiral (t = 37.4 s), growth of a second turn in the spiral (t = 152.2 s) and filament rearrangements (t = 164.9 s).

(legend continued on next page)

stiffness of the Snf7 filaments, which is characterized by its persistence length ℓ_p. We estimated this persistence length from the amplitude of the thermal fluctuations of isolated Snf7 filaments on a supported bilayer observed by HS-AFM. We obtained $\ell_p = 260$ nm (Supplemental Information, Supplemental Mathematical Modeling part 2). This is in order-of-magnitude agreement with an estimate from a numerical model of Snf7 flexion $(\ell_p \approx 800$ nm$)$ (Shen et al., 2014). This value is higher than for DNA $(\ell_p = 50$ nm$)$, but smaller than for cytoskeletal filaments $(\ell_p = 15$ μm for actin and $\ell_p = 6$ mm for microtubules) (Howard, 2001), implying that they are intrinsically soft enough to be deformed by moderate forces.

In our theoretical 2D spring model, spirals become significantly deformed when μ exceeds a threshold surface energy $\mu^* = 1/(2 - \pi/\sqrt{3})(k_B T \ell_p / R^2 b)\log(R/b)$ (where k_B is the Boltzmann constant and T the temperature, see Supplemental Information, Supplemental Mathematical Modeling part 3). Our experimental estimates for R ($\sim$130 nm), b (17 nm) and ℓ_p (260 nm) implied $\mu^* = 4.0 \times 10^{-5}$ J·m^{-2}, 8 times smaller than the measured experimental μ. It indicates that the Snf7 polymerization energy, even if underestimated, is sufficient to induce strong deformations. Our data show that the Snf7 spirals can deform as spiral springs, and can self-load through a mechanism where deformation is mostly generated by growth of a filament.

Snf7 Spirals Expansion Leads to Membrane Deformation

Having established that the Snf7 spirals can deform to store significant elastic energy, we wondered if they could release this energy to deform the membrane. After several hours of Snf7 incubation, holes (called "pores" in the following) spontaneously appeared in a few GUVs, releasing membrane tension. Surprisingly, instead of bursting, the GUV membrane shrunk from the rim of the pore toward the opposite side of the GUV (Movie S7). Occasionally, this process stopped before the vesicle had fully collapsed, and stable vesicles with open pores were observed (Figure 6A; Movie S8). In this case, a stronger fluorescence signal is seen at the rim.

To understand the stronger signal of the membrane marker at the rim, we imaged these opened GUVs by thin-section EM: the membrane at the rim of the pore was rolled toward the interior of the vesicle (Figure 6B). This process is known as curling, and has previously been observed in a number of situations, including polymersomes (Mabrouk et al., 2009) and during the bursting of red blood cells (Callan-Jones et al., 2012). It occurs when an area difference appears between the two sides of a bi-layered surface. We hypothesized that curling could be driven by the expansion of the previously compressed Snf7 layer following pore formation. To quantitatively study the plausibility of such scenario, we used our theoretical model to compute the ratio

of the surface occupied by a compressed spiral ($A_{compressed}$) to that of the a relaxed state ($A_{relaxed}$) (see Supplemental Information, Supplemental Mathematical Modeling part 3) as

$$\frac{A_{compressed}}{A_{relaxed}} = \left[\frac{\pi}{2\sqrt{3}} + \left(1 - \frac{\pi}{2\sqrt{3}} \right) \sqrt{\frac{\mu^*}{\mu}} \right]^2 = 94\% \qquad (2)$$

The value of this ratio implies that, during stress release, the Snf7 layer would expand by 6% relative to the underlying, almost inextensible lipid bilayer (Figure 6C). As a result, the membrane of the GUV would curl inward (Figure 6D), consistently with our observations in fluorescence and electron microscopy. The preferred curvature of the curl can be estimated as $r_c = (d/2)(A_{relaxed} + A_{compressed})/(A_{relaxed} - A_{compressed})$, where $2d$ is the total thickness of the lipid-Snf7 sandwich. Using AFM to measure the thickness of the Snf7 coated membranes, we estimated $2d = 9$ nm (5 nm for the membrane plus 4 nm for the Snf7 coat). With this value, we can estimate $r_c = 37$ nm. Considering previous studies (Callan-Jones et al., 2012), r_c corresponds to the curvature of the innermost roll observed (Figure 6D). Experimentally, we find a mean radius of $r_c = 39 \pm 6$ nm (n = 9). Therefore curling in the opened vesicles can be explained by the expansion of the Snf7 spiral springs. During this expansion, the spirals release their compression energy accumulated during polymerization.

We wondered how much energy was stored in a single spiral, as compared to the energy required for budding of a vesicle. Our elastic model implies that the lateral compression of a single Snf7 spiral corresponds to the accumulation of an elastic energy $\Delta E = 170 \, k_B T = 7.0 \times 10^{-19}$ J, which is bigger than the bending energy $4\pi\kappa = 160 \, k_B T$ required to form a spherical membrane bud (see Supplemental Information, Supplemental Mathematical Modeling). Thus, a single spiral can accumulate enough elastic energy to form a spherical bud when released.

In summary, we show here that Snf7 filaments display the ability to act as spiral springs that load through polymerization. The release of the compression stress accumulated during the deformation of the Snf7 spiral is sufficient to drive membrane deformation.

DISCUSSION

In this study, we first showed that lipid membranes trigger the formation of wild-type Snf7 assemblies at their surface through a process of nucleation-growth. The patch nucleation rate is low (less than 1 seed.μm^{-2}.hour^{-1}) which explains the necessity of a polymerization-activated mutant to observe the same assemblies by EM in previous studies (Henne et al., 2012). We found that the circular arrays formed by Snf7 on these membranes are spirals made of a single filament looping and interacting onto itself. This confirms that the interaction with membrane in vitro retains the

(F) HS-AFM time-lapse sequence of an isolated Snf7 spiral. Arrowheads show: growth of the spiral at the two-turn stage (t = 67.2 s), and filament split (t = 75.7 s) leading to the three turns stage.

(G) The equatorial kymograph (yellow rectangle) of this growing spiral: the innermost turn radius decreases from 22 nm to 14 nm upon formation of the third turn.

(H) Dynamics of filament splitting and fusing in two Snf7 spirals (rows) observed by HS-AFM. Arrowheads show displacement of the splitting points.

(I) Time plot of the outer radius of five growing Snf7 spirals followed by HS-AFM. The origin of all curves is the apparition of the first turn. The thick curve is the average of all curves. [Snf7] = 1 μM. See also Figure S4.
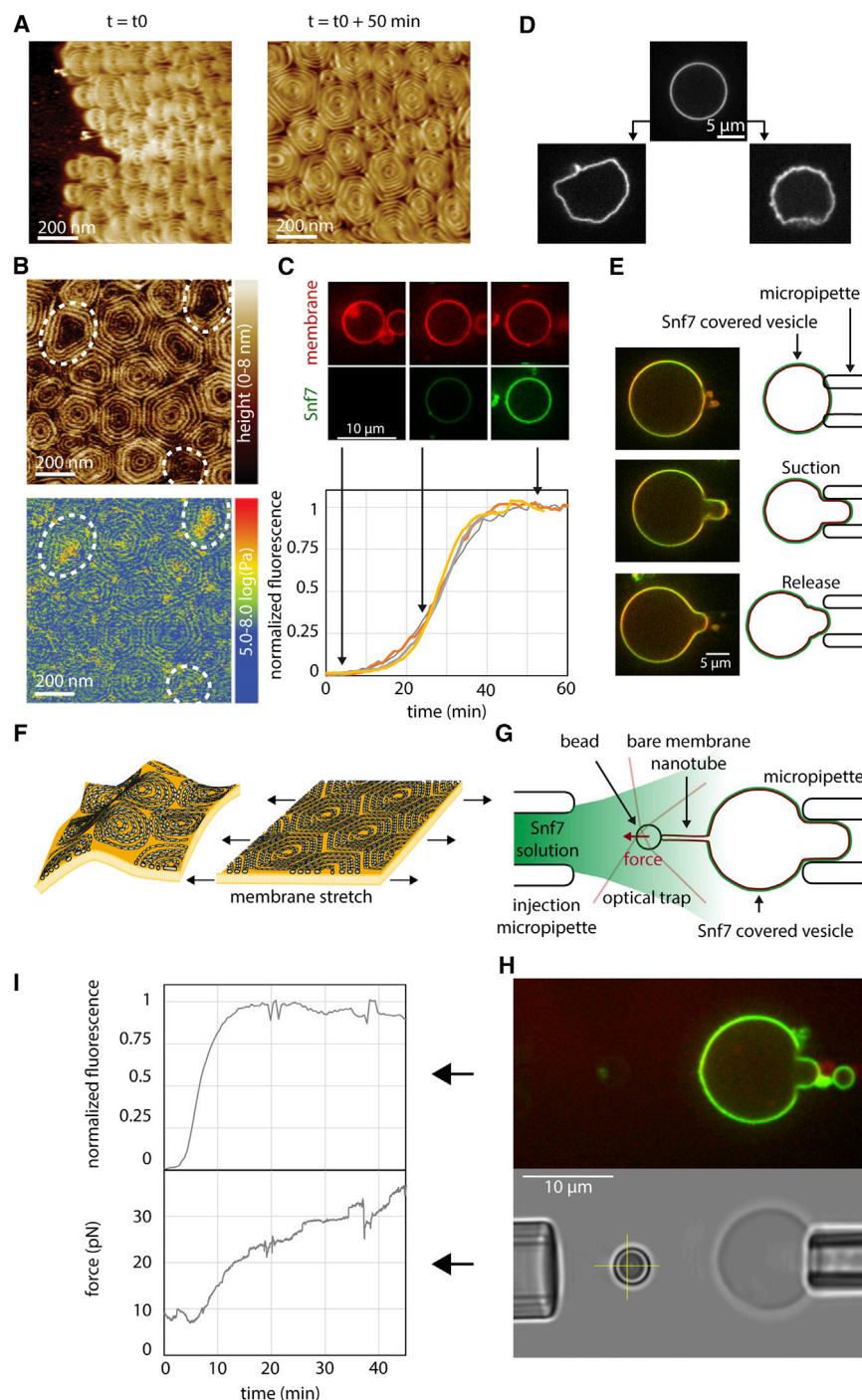
Figure 5. Build-up of Lateral Compression in Snf7 Spirals by Polymerization

(A) HS-AFM images of Snf7 spirals acquiring polygonal shapes with time.

(B) AFM Topography and nanomechanical mapping of polygonal Snf7 spirals. A significant proportion of spirals (dashed outlines) have a lower center with increased mechanical stiffness.

(C) Snf7 polymerization on GUVs made of DOPC 60% / DOPS 40% + Rhodamine-PE 0.1% (red), 0.003% DOPE-Peg2000-Biotin. GUVs are incubated with 500 nM Snf7-Alexa488 (green). Top: SDC images of a GUV equatorial plane during Snf7 polymerization. Bottom: fluorescence intensity (equatorial plane) of 4 GUVs with time.

(D) GUVs before (top) and after (bottom) several hours of incubation with Snf7-Alexa488.

(E) Snf7 coated GUVs keep the aspirated shape after release from the micropipette.

(F) Sketch of membrane stretching by Snf7 spiral compression.

(G) Schematic of the membrane tension measurement setup combining holding pipette, injection pipette, bead within an optical trap, giant vesicle (red) and Snf7 (green).

(H) Top image: SDC image of a membrane tension measurement experiment (red = membrane, green = Snf7-Alexa488). Note that Snf7-Alexa488 did not polymerize on the membrane nanotube. Bottom: brightfield image of the same vesicle. The yellow cross indicates the resting position of the bead held by the optical trap.

(I) Top: Normalized Snf7 fluorescence intensity versus time (measured from equatorial plane); bottom: force exerted by the membrane nanotube on the bead versus time. See also Figure S5.

rals, because of the relatively high flexibility of the Snf7 filament can be deformed by lateral compression. Moreover, we show that the expansion of compressed spirals can lead to membrane deformation if confinement is released (Figure 6). These observations imply that Snf7 can work as a two-dimensional spring, being able to compress and expand. In the following, we discuss how this spring-like activity highlighted by our study is relevant for the in vivo situation.

In vivo, it is unlikely that a densely packed array of ESCRT-III spirals is present at the surface of membranes, which

spiral structure recently observed in solution (Shen et al., 2014) or in vivo (Cashikar et al., 2014; Hanson et al., 2008). We also find that Snf7 filaments can bundle into double-stranded filaments, probably through parallel lateral interactions.

In our assay, spirals become tightly packed into polygonal lattice at the surface of the supported bilayers. The packing of these spirals is correlated with the increase of lateral compression within the Snf7 coat. These data implies that ESCRT-III spi-

may question the physiological relevance of the spiral compression observed in large patches of ESCRT-III. But the confinement required for such lateral compression might come from other membrane proteins, which may provide walls into which single spirals could be confined. In the membrane of MVBs, Lamp1 and 2 are particularly enriched (Bissig and Gruenberg, 2014) and may provide a scaffold onto which ESCRT-III spirals could be compressed. Of course, compression being isotropic
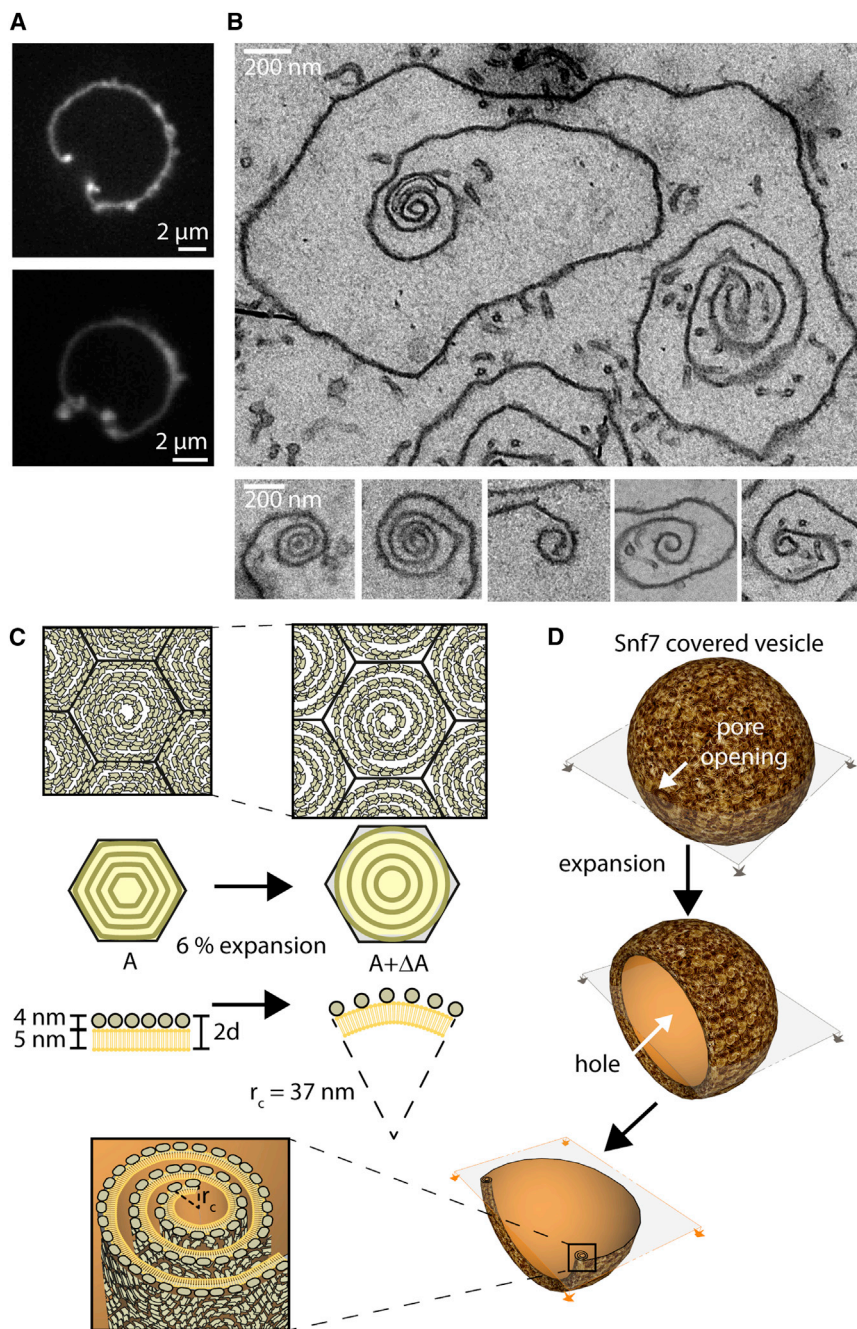
(A) Confocal sections of Snf7 coated vesicles displaying stable holes. Fluorescence is more intense at the rim of the pore.

(B) EM thin section image of a Snf7 coated vesicle with a stable pore. Note the curling of the membrane rim. Several other examples of membrane curling are shown in lower panels.

(C) Sketch of the expected curvature generated by expansion of compressed Snf7 spirals.

(D) Sketch of the pore opening and curling of Snf7 coated vesicle. Bottom images show the expected section of a stable pore in the GUV and a zoom on the membrane curled region.

ing the inner filaments. Accordingly, we find that even for non-laterally constrained spirals, the inner turn of the spirals tightens when the number of turns in the spirals goes above three.

Whatever the source of compression is, our observations show the ability of Snf7 spirals to deform elastically and accumulate potential energy that can be used for membrane deformation. But how would such energy drive membrane budding? It was previously proposed that the polymerization of ESCRT-III could enclose a patch of membrane and then, by reduction of the length of the Snf7 rim, the membrane would be folded into a bud in the middle of the Snf7 polymer. The ESCRT-III rim reduction has been proposed to be mediated by depolymerization of the Snf7 spiral (lasso model, (Saksena et al., 2009)) or by further inward polymerization of the Snf7 spiral (Cashikar et al., 2014). Our data suggests that Snf7 spiral spring could mediate the rim reduction by its elastic compression down to a 14 nm radius. However in this scenario, because the membrane is fluid, it is difficult to picture how the force of the spiral spring would be transmitted to the membrane. We propose that cargoes play an essential role for the force transmission (Figure 7,left): the rim reduction would lead to compaction of enclosed membrane cargoes. Theoretical (Derganc et al., 2013) and experimental studies (Stachowiak et al., 2012) indicate that highly dense cargoes could mediate budding by asymmetric crowding. This is consistent with a recent in vivo study (Mageswaran et al., 2015) where ILVs budding was critically dependent on accumulation of cargoes within ESCRT-III assemblies.

in this in vivo case, compressed spirals would stay circular, instead of polygonal.

But another source of lateral compression is intrinsic to the spiral structure, and is present in single Snf7 spiral even in the absence of any external confining structures. We show that the filaments curl spontaneously at 20–30 nm, implying that if they grow at a different radius, they are under mechanical stress. Indeed, when spirals are broken, all the pieces of filaments further curl to a smaller radius. Thus, in the spiral structure, filaments with a radius larger than 25 nm are stretched, compress-

Another possibility is that the out-of-plane buckling of the spring itself might drive the invagination of the membrane (Figure 7, right) (Lenz et al., 2009), which could explain the formation
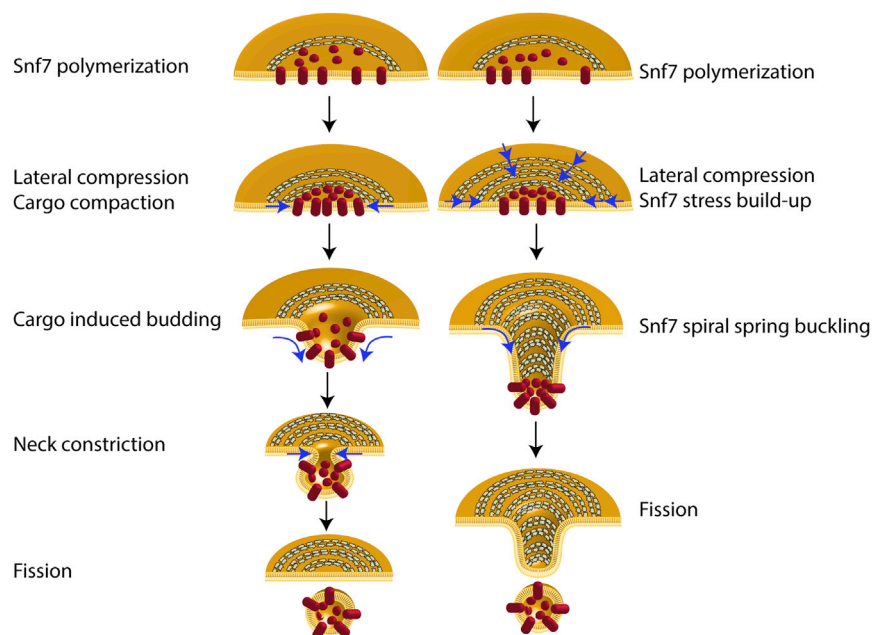
Figure 7. Models of ESCRT-III Mediated Budding and Fission of Intra-lumenal Vesicles

Left: cargo sequestration and ESCRT-III lateral compression induces membrane budding. Further ESCRT-III narrowing might lead to fission. Right: ESCRT-III lateral compression leads to buckling.

(Figure labels — left column: Snf7 polymerization; Lateral compression, Cargo compaction; Cargo induced budding; Neck constriction; Fission. Right column: Snf7 polymerization; Lateral compression, Snf7 stress build-up; Snf7 spiral spring buckling; Fission.)

β-mercaptoethanol [pH 7.5]) and Vps20 (10 μM stock solution in 20 mM HEPES 100 mM NaCl [pH 7.5]) were kept unlabeled.

Giant Unilamellar Vesicles and Large Unilamellar Vesicles Preparation

GUVs were prepared by electroformation using DOPC and DOPS mixtures, purchased from Avanti Polar Lipids (Alabaster, USA). When necessary, 0.1% fluorescent lipids (1,2-dioleoyl-sn-glycero-3-phosphoethanolamine-N-(lissamine rhodamine B sulfonyl)) (Rhodamine-PE) were added.

LUVs were prepared by evaporating in a round-bottom glass tube, a volume of lipid mix (DOPC: DOPS, 6:4, mol:mol) containing 1 mg of total lipids. After addition of 200 μl of buffer, the tube was vortexed and freeze-thaw 3 times. This solution is kept at −20°C until use.

Unless otherwise noted, the buffer used for all experiments is composed of 20 mM Tris HCl (pH 6.8), 200 mM NaCl, 1 mM MgCl$_2$.

Optical Microscopy of Membrane Assays

For confocal and TIRF imaging, a coverslip is cleaned with water and ethanol, and then plasma-cleaned for 2 min (PDC-32G, Harrick Plasma, NY, USA). The coverslip is assembled to a flow chamber (sticky-Slide VI 0.4, Ibidi, Munich, Germany), with one entry connected to a syringe pump (Aladdin, World Precision Instruments, Sarasota, FL, USA), and the other left open for sequential introduction of other solutions. The flow chamber is initially filled by 200 μl of buffer. 5 μl of GUVs are flushed in the flow chamber (see Extended Experimental Procedures for methods to get supported membranes or partially adhered vesicles).

Imaging is performed using an inverted microscope assembled by 3i (Intelligent Imaging Innovation, Denver, USA) and Nikon (Eclipse C1, Nikon, Tokyo, Japan). For SDC imaging, a 2-μm-thick volume stack (1 μm above and below the supported membrane) is acquired then rendered to 2D by maximum intensity projection. TIRF Imaging is performed using a motorized Nikon TIRF system. The number of molecules within Snf7 oligomers is estimated by calibrating the microscope with commercially available fluorescent DNA origamis (GATTA-Brightness 9R and 18R, GATTAquant, Braunschweig, Germany) (Figure S6).

Optical Tweezers Tube Pulling Experiment

A modified version of a published setup (Morlot et al., 2012) allows simultaneous brightfield imaging, SDC microscopy, and optical tweezing on an inverted Nikon eclipse Ti microscope. A GUV is aspirated within a micropipette connected to a motorized micromanipulator (MP-285, Sutter Instrument, Novato, CA, USA) and a pressure control system (MFCS-VAC −69 mbar, Fluigent, Villejuif, France) that sets the aspiration pressure ΔP. A membrane nanotube is then pulled out from the GUV through a streptavidin-coated bead (3.05 μm diameter, Spherotec, Lake Forest, IL, USA) held in a fixed optical trap. The optical trap was custom-made with a continuous 5 W 1064 nm fiber laser (ML5-CW-P-TKS-OTS, Manlight, Lannion, France) focused through a 100X 1.3 NA oil immersion objective. The force F exerted on the bead was calculated from the Hooke's law: F = k.Δx, where k is the stiffness of the trap (k = 60 pN.μm^{-1}) and Δx the displacement of the bead from its

of membrane tubules by overexpression of the human homolog of Snf7 (Hanson et al., 2008). This model implies that the flat membrane could be a metastable state of this elastic system: the stored elastic energy could be suddenly released upon external activation. Our observation that the spirals adopt a curved inverted dome shape (Figure 5B) is consistent with this model, and an estimate of the energy stored in one spiral further confirm that this energy is sufficient to bud the membrane into a sphere.

The spring-like properties of Snf7 filaments also nourish our understanding of ESCRT-III role in membrane fission. Because of their high flexibility, Snf7 filaments can grow at radii different from their preferred radius of curvature if steric or mechanical constraints force them to do so. This feature explains how Snf7 filaments could adapt to the wide range of radii observed in the various ESCRT-III mediated fission reactions: from microns in abscission and hundreds of nanometer in virus budding, down to tens of nanometers in ILV formation and membrane repair (Jimenez et al., 2014). However, the smallest size of the inner turn is on average 18 nm radius, which is far from the 1.4 nm observed with dynamin to finalize fission (Sundborger et al., 2014). This raises questions regarding the mechanism necessary to provide fission and pore closure and supports the role of other ESCRT-III proteins and lipids in these reactions.

EXPERIMENTAL PROCEDURES

Protein Purification and Labeling

Snf7 (Addgene plasmid no. 21492), Escrt-II (Addgene plasmid no. 17633) and Vps20 (Addgene plasmid no. 21490) were purified as previously described (Hierro et al., 2004; Wollert et al., 2009). Snf7 stock solution was 2.5 μM in 20 mM HEPES, 100 mM NaCl (pH 8). Snf7 was labeled either with TFP-Alexa-488 (Life technologie product no. A37563) or with NHS-Atto 647N (Atto-tec product no. AD 647N-3). Escrt-II (20 μM stock solution in 50 mM Tris, 150 mM NaCl, 5 mM

equilibrium position. Snf7 was injected close to the nanotube with a second micropipette connected to another channel of the Fluigent pressure control system.

Electron Microscopy

For negative stain EM observations, LUVs were incubated with Snf7 in suspension, spun down (4 min at 4,000 g), washed and then adsorbed onto glow-discharged Formvar coated EM grids. The samples were negatively stained for 30 s with 2% uranyl acetate before visualization.

Ultrathin sectioning of Snf7 bound LUVs fixed in epon was performed using a microtome (Leica Ultracut) at a cutting angle of 6°. Sections were put on glow-discharged carbon-coated formvar grids and imaged with a Tecnai G2 Sphera (FEI) electron microscope.

AFM and HS-AFM

For both PF-QNM AFM and HS-AFM, GUVs, prepared as described above, were adsorbed to the mica support followed by protein addition. For PF-QNM AFM experiments 5 μl of the GUVs and for HS-AFM experiments 0.5 μl of GUVs were deposited onto freshly cleaved mica supports pre-incubated with adsorption buffer (220 mM NaCl, 10 mM HEPES, 2 mM MgCl₂, [pH 7.4]). Supported lipid bilayers were first imaged to assess the quality of the lipid bilayer preparation before injecting Snf7 into the fluid cell to a concentration of ~500 nM. Formation of Snf7 assemblies were observable ~30 min after Snf7 injection.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and eight movies and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.017.

REFERENCES

Adell, M.A.Y., Vogel, G.F., Pakdel, M., Müller, M., Lindner, H., Hess, M.W., and Teis, D. (2014). Coordinated binding of Vps4 to ESCRT-III drives membrane neck constriction during MVB vesicle formation. J. Cell Biol. 205, 33–49.

Babst, M., Katzmann, D.J., Estepa-Sabal, E.J., Meerloo, T., and Emr, S.D. (2002). Escrt-III: an endosome-associated heterooligomeric protein complex required for mvb sorting. Dev. Cell 3, 271–282.

Bissig, C., and Gruenberg, J. (2014). ALIX and the multivesicular endosome: ALIX in Wonderland. Trends Cell Biol. 24, 19–25.

Buchkovich, N.J., Henne, W.M., Tang, S., and Emr, S.D. (2013). Essential N-terminal insertion motif anchors the ESCRT-III filament during MVB vesicle formation. Dev. Cell 27, 201–214.

Callan-Jones, A., Albarran Arriagada, O.E., Massiera, G., Lorman, V., and Abkarian, M. (2012). Red blood cell membrane dynamics during malaria parasite egress. Biophys. J. 103, 2475–2483.

Carlton, J.G., Agromayor, M., and Martin-Serrano, J. (2008). Differential requirements for Alix and ESCRT-III in cytokinesis and HIV-1 release. Proc. Natl. Acad. Sci. USA 105, 10541–10546.

Cashikar, A.G., Shim, S., Roth, R., Maldazys, M.R., Heuser, J.E., and Hanson, P.I. (2014). Structure of cellular ESCRT-III spirals and their relationship to HIV budding. eLife 3, e02184.

Casuso, I., Khao, J., Chami, M., Paul-Gilloteaux, P., Husain, M., Duneau, J.-P., Stahlberg, H., Sturgis, J.N., and Scheuring, S. (2012). Characterization of the motion of membrane proteins using high-speed atomic force microscopy. Nat. Nanotechnol. 7, 525–529.

Coonrod, E.M., and Stevens, T.H. (2010). The yeast vps class E mutants: the beginning of the molecular genetic analysis of multivesicular body biogenesis. Mol. Biol. Cell 21, 4057–4060.

Cuvelier, D., Derényi, I., Bassereau, P., and Nassoy, P. (2005). Coalescence of membrane tethers: experiments, theory, and applications. Biophys. J. 88, 2714–2726.

Derganc, J., Antonny, B., and Copič, A. (2013). Membrane bending: the power of protein imbalance. Trends Biochem. Sci. 38, 576–584.

Elia, N., Sougrat, R., Spurlin, T.A., Hurley, J.H., and Lippincott-Schwartz, J. (2011). Dynamics of endosomal sorting complex required for transport (ESCRT) machinery during cytokinesis and its role in abscission. Proc. Natl. Acad. Sci. USA 108, 4846–4851.

Fabrikant, G., Lata, S., Riches, J.D., Briggs, J.A., Weissenhorn, W., and Kozlov, M.M. (2009). Computational model of membrane fission catalyzed by ESCRT-III. PLoS Comput. Biol. 5, e1000575.

Guizetti, J., Schermelleh, L., Mäntler, J., Maar, S., Poser, I., Leonhardt, H., Müller-Reichert, T., and Gerlich, D.W. (2011). Cortical constriction during abscission involves helices of ESCRT-III-dependent filaments. Science 331, 1616–1620.

Hanson, P.I., Roth, R., Lin, Y., and Heuser, J.E. (2008). Plasma membrane deformation by circular arrays of ESCRT-III protein filaments. J. Cell Biol. 180, 389–402.

Henne, W.M., Buchkovich, N.J., Zhao, Y., and Emr, S.D. (2012). The endosomal sorting complex ESCRT-II mediates the assembly and architecture of ESCRT-III helices. Cell 151, 356–371.

Hierro, A., Sun, J., Rusnak, A.S., Kim, J., Prag, G., Emr, S.D., and Hurley, J.H. (2004). Structure of the ESCRT-II endosomal trafficking complex. Nature 431, 221–225.

Howard, J. (2001). Mechanics of Motor Proteins and the Cytoskeleton (Sunderland, Massachusetts: Sinauer Associates).

Jimenez, A.J., Maiuri, P., Lafaurie-Janvore, J., Divoux, S., Piel, M., and Perez, F. (2014). ESCRT machinery is required for plasma membrane repair. Science 343, 1247136.

Lata, S., Schoehn, G., Jain, A., Pires, R., Piehler, J., Gottlinger, H.G., and Weissenhorn, W. (2008). Helical structures of ESCRT-III are disassembled by VPS4. Science 321, 1354–1357.

Lenz, M., Crow, D.J., and Joanny, J.-F. (2009). Membrane buckling induced by curved filaments. Phys. Rev. Lett. *103*, 038101.

Mabrouk, E., Cuvelier, D., Brochard-Wyart, F., Nassoy, P., and Li, M.-H. (2009). Bursting of sensitive polymersomes induced by curling. Proc. Natl. Acad. Sci. USA *106*, 7294–7298.

Mageswaran, S.K., Johnson, N.K., Odorizzi, G., and Babst, M. (2015). Constitutively active ESCRT-II suppresses the MVB-sorting phenotype of ESCRT-0 and ESCRT-I mutants. Mol. Biol. Cell *26*, 554–568.

McCullough, J., Colf, L.A., and Sundquist, W.I. (2013). Membrane fission reactions of the mammalian ESCRT pathway. Annu. Rev. Biochem. *82*, 663–692.

Morlot, S., Galli, V., Klein, M., Chiaruttini, N., Manzi, J., Humbert, F., Dinis, L., Lenz, M., Cappello, G., and Roux, A. (2012). Membrane shape at the edge of the dynamin helix sets location and duration of the fission reaction. Cell *151*, 619–629.

Pires, R., Hartlieb, B., Signor, L., Schoehn, G., Lata, S., Roessle, M., Moriscot, C., Popov, S., Hinz, A., Jamin, M., et al. (2009). A crescent-shaped ALIX dimer targets ESCRT-III CHMP4 filaments. Structure *17*, 843–856.

Saksena, S., Wahlman, J., Teis, D., Johnson, A.E., and Emr, S.D. (2009). Functional reconstitution of ESCRT-III assembly and disassembly. Cell *136*, 97–109.

Saleem, M., Morlot, S., Hohendahl, A., Manzi, J., Lenz, M., and Roux, A. (2015). A balance between membrane elasticity and polymerization energy sets the shape of spherical clathrin coats. Nat. Commun. *6*, 6249.

Scheuring, S., Seguin, J., Marco, S., Lévy, D., Robert, B., and Rigaud, J.-L. (2003). Nanodissection and high-resolution imaging of the Rhodopseudomonas viridis photosynthetic core complex in native membranes by AFM. Atomic force microscopy. Proc. Natl. Acad. Sci. USA *100*, 1690–1693.

Shen, Q.-T., Schuh, A.L., Zheng, Y., Quinney, K., Wang, L., Hanna, M., Mitchell, J.C., Otegui, M.S., Ahlquist, P., Cui, Q., and Audhya, A. (2014). Structural analysis and modeling reveals new mechanisms governing ESCRT-III spiral filament assembly. J. Cell Biol. *206*, 763–777.

Stachowiak, J.C., Schmid, E.M., Ryan, C.J., Ann, H.S., Sasaki, D.Y., Sherman, M.B., Geissler, P.L., Fletcher, D.A., and Hayden, C.C. (2012). Membrane bending by protein-protein crowding. Nat. Cell Biol. *14*, 944–949.

Sundborger, A.C., Fang, S., Heymann, J.A., Ray, P., Chappie, J.S., and Hinshaw, J.E. (2014). A dynamin mutant defines a superconstricted prefission state. Cell Rep. *8*, 734–742.

von Schwedler, U.K., Stuchell, M., Müller, B., Ward, D.M., Chung, H.-Y., Morita, E., Wang, H.E., Davis, T., He, G.-P., Cimbora, D.M., et al. (2003). The protein network of HIV budding. Cell *114*, 701–713.

Wollert, T., and Hurley, J.H. (2010). Molecular mechanism of multivesicular body biogenesis by ESCRT complexes. Nature *464*, 864–869.

Wollert, T., Wunder, C., Lippincott-Schwartz, J., and Hurley, J.H. (2009). Membrane scission by the ESCRT-III complex. Nature *458*, 172–177.

Cell

53BP1 and the LINC Complex Promote Microtubule-Dependent DSB Mobility and DNA Repair

Graphical Abstract
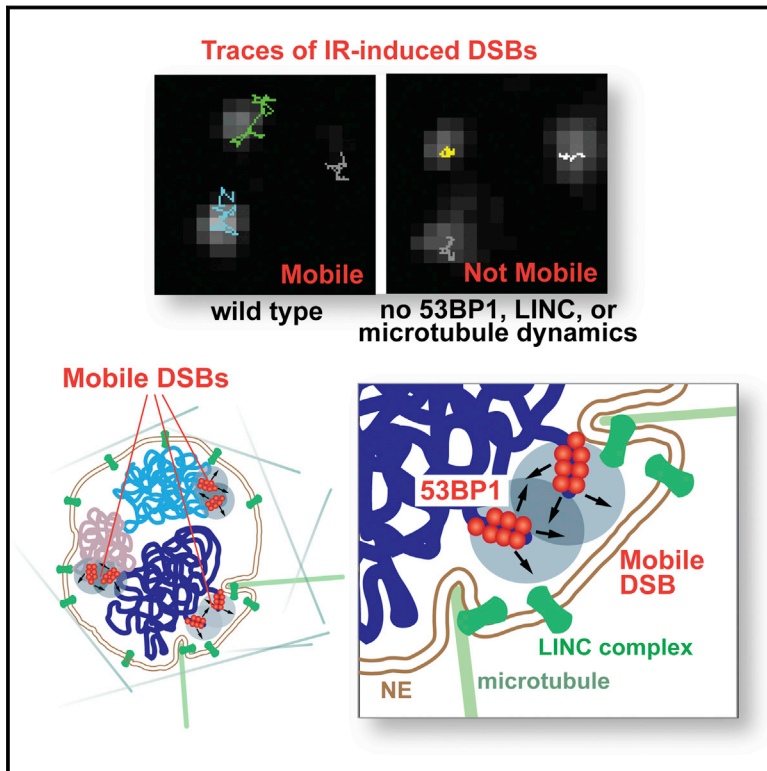
Authors

Francisca Lottersberger,
Roos Anna Karssemeijer,
Nadya Dimitrova, Titia de Lange

Correspondence

delange@rockefeller.edu

In Brief

Increased chromatin mobility at sites of double-stranded breaks is mediated by 53BP1, SUN1/2 in the nuclear envelope, and microtubule dynamics and contributes to aberrant DNA repair in cells with multiple DNA breaks.

Highlights

- Dysfunctional telomeres and DSBs are more mobile than undamaged chromatin

- 53BP1, the LINC complex, and dynamic microtubules are required for DSB mobility

- DSB mobility promotes radial formation in PARPi-treated BRCA1-deficient cells

- DSB mobility is proposed to represent an error correction mechanism

CellPress

53BP1 and the LINC Complex Promote Microtubule-Dependent DSB Mobility and DNA Repair

Francisca Lottersberger,[1] Roos Anna Karssemeijer,[1] Nadya Dimitrova,[1,2] and Titia de Lange[1,*]

[1]Laboratory for Cell Biology and Genetics, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA
[2]Present address: Department for Molecular, Cellular, and Developmental Biology, Yale University, 219 Prospect Street, New Haven, CT 06511, USA
*Correspondence: delange@rockefeller.edu
http://dx.doi.org/10.1016/j.cell.2015.09.057

SUMMARY

Increased mobility of chromatin surrounding double-strand breaks (DSBs) has been noted in yeast and mammalian cells but the underlying mechanism and its contribution to DSB repair remain unclear. Here, we use a telomere-based system to track DNA damage foci with high resolution in living cells. We find that the greater mobility of damaged chromatin requires 53BP1, SUN1/2 in the linker of the nucleoskeleton, and cytoskeleton (LINC) complex and dynamic microtubules. The data further demonstrate that the excursions promote non-homologous end joining of dysfunctional telomeres and implicated Nesprin-4 and kinesins in telomere fusion. 53BP1/LINC/microtubule-dependent mobility is also evident at irradiation-induced DSBs and contributes to the mis-rejoining of drug-induced DSBs in BRCA1-deficient cells showing that DSB mobility can be detrimental in cells with numerous DSBs. In contrast, under physiological conditions where cells have only one or a few lesions, DSB mobility is proposed to prevent errors in DNA repair.

INTRODUCTION

The integrity of eukaryotic genomes is perpetually threatened by the formation of double-stranded breaks (DSBs), which can arise due to errors in DNA metabolism or genotoxic insults, such as chemotherapeutic agents. The repair of DSBs is a critical aspect of genome maintenance, despite the fact that non-cycling cells experience only a few DSBs per day (Fumagalli et al., 2012; U. Herbig, personal communication). In G1, DSBs are repaired by non-homologous end-joining (NHEJ) whereas replicating cells can also use a second pathway, homology-directed repair (HDR), to restore genome integrity. NHEJ and HDR are highly regulated to avoid ectopic repair, which can generate translocations, multicentric chromosomes, and other deleterious chromosome rearrangements.

The role of the DNA damage response factor 53BP1 in DSB repair and its contribution to cell-cycle appropriate execution of NHEJ and HDR has been studied extensively (reviewed in Escribano-Díaz et al., 2013; Panier and Boulton, 2014; Zimmermann and de Lange, 2014). 53BP1 accumulates at sites of DNA damage through a dual interaction between its Tudor domain with constitutively dimethylated histone H4 (H4K20diMe) and its UDR domain with ubiquitylated histone H2A (H2AK15Ub), which marks sites of DNA damage. Many of the functions of 53BP1 are mediated by binding partners that associate with the 53BP1 N terminus upon phosphorylation of ST/Q sites by the ataxia telangiectasia mutated (ATM) and ataxia telangiectasia and Rad3 related (ATR) kinases.

A critical role of 53BP1 is to limit the 5′ resection of the broken ends in a cell-cycle-dependent manner. Whereas inappropriate resection in G1 will impede the repair of DSBs by NHEJ, resection is needed for HDR in S/G2. Inhibition of 5′ end resection in G1 is primarily mediated by the 53BP1-bound Rif1 and Rev7/MAD2L2, but the mechanism by which resection is blocked is unknown (Chapman et al., 2013; Di Virgilio et al., 2013; Escribano-Díaz et al., 2013; Feng et al., 2013; Zimmermann et al., 2013; Boersma et al., 2015; Xu et al., 2015). In S/G2, the action of Rif1 and Rev7/MAD2L2 are counteracted by BRCA1, allowing resection and generating the 3′ overhangs required for HDR. A second 53BP1-interacting factor, PTIP, has an auxiliary role that involves end trimming by the Artemis nuclease (Munoz et al., 2007; Callen et al., 2013; Wang et al., 2014).

The contribution of 53BP1 to DSB repair pathway choice has received considerable attention in the context of the treatment of BRCA1-deficient cancers with poly(ADP-ribose) polymerase inhibitors (PARPi) (reviewed in Banerjee et al., 2010). PARP inhibition results in a large number of persistent single-stranded (ss) gaps that are converted into DSBs by DNA replication. In absence of BRCA1, the inefficiency of 5′ end resection allows NHEJ to dominate the repair. When many broken ends persist, NHEJ can promote mis-rejoining of broken chromatids, forming radial chromosomes and chromosome aberrations that have lethal consequences. This mis-repair of DSBs determines the synthetic lethality of PARP inhibition and HR deficiency. Removal of 53BP1 in this setting blocks the formation of mis-repaired chromosomes, in part by alleviating the inhibition of resection and hence restoring HDR (Cao et al., 2009; Bouwman et al., 2010; Bunting et al., 2010; Chapman et al., 2013; Di Virgilio et al., 2013; Zimmermann et al., 2013; Xu et al., 2015). Indeed, absence of Rif1 or MAD2L2 also minimizes the formation of mis-repaired chromosomes in PARPi-treated BRCA1-negative cells. However, 53BP1 has a greater effect than Rif1 (Zimmermann et al.,

2013), suggesting a second mechanism by which 53BP1 promotes mis-rejoining.

We have used dysfunctional telomeres to investigate the second, Rif1-independent function of 53BP1. Mammalian telomeres are protected from the DNA damage response (DDR) by the six-member shelterin protein complex residing on the telomeric TTAGGG repeats (reviewed in Palm and de Lange, 2008). Removal of TRF2 from shelterin unleashes two pathways that normally are repressed at telomeres. Telomeres lacking TRF2 activate ATM kinase signaling, leading to Chk2 phosphorylation and the accumulation of 53BP1 at telomeres. In addition, TRF2 loss from telomeres renders them highly susceptible to Ku70/80- and DNA ligase IV (lig4)-dependent classical(c)-NHEJ.

In addition to blocking resection at dysfunctional telomeres, 53BP1 alters their mobility. After loss of TRF2, telomeres travel greater distances and roam larger subnuclear territories than functional telomeres (Dimitrova et al., 2008). This effect was also observed upon telomere deprotection with a TIN2 short hairpin RNA (shRNA) (Chen et al., 2013). The altered mobility of dysfunctional telomeres is strictly dependent on 53BP1 but not influenced by Rif1 (Zimmermann et al., 2013) or Rev7/MAD2L2 (Boersma et al., 2015). Given that, in G1, the fusion of two telomeres involves chromosome ends that are spatially separated, we speculated that 53BP1-dependent mobility could stimulate c-NHEJ by increasing the chance that two ends become juxtaposed. Indeed, 53BP1 is required for telomere-telomere fusions (Dimitrova et al., 2008) and this dependency cannot be fully explained by the ability of 53BP1 to block resection (Zimmermann et al., 2013).

In budding yeast, increased chromatin mobility occurs near an I-Sce-induced DSB and, to lesser extent, at the level of global chromatin (Dion et al., 2012; Miné-Hattab and Rothstein, 2012; Seeber et al., 2013), possibly enhancing the homology search needed for HDR (Agmon et al., 2013). Similarly, in fission yeast, DSBs associate with the LINC complex in a process that promotes HDR (Swartz et al., 2014). However, the data on the mobility of DSBs in mammalian cells has been equivocal (reviewed in Dion and Gasser, 2013). Ionizing radiation (IR)-induced DSBs show an ATM-dependent increase in mobility (Neumaier et al., 2012; Becker et al., 2014), lesions induced by α-particles or I-PpoI have been inferred to move (Aten et al., 2004; Falk et al., 2007; Gandhi et al., 2012), and directed movement occurs during telomere recombination in the context of the alternative lengthening of telomeres (ALT) pathway (Cho et al., 2014). However, other findings have argued against an altered mobility of DSBs (Kruhlak et al., 2006; Soutoglou et al., 2007; Jakob et al., 2009).

Using time-lapse imaging of conditional TRF2 knockout (KO) mouse embryonic fibroblasts (MEFs) as a model system, we demonstrate here that 53BP1-dependent chromatin mobility is mediated by microtubules and the LINC complex. The LINC complex spans the inner and outer membranes (INM and ONM, respectively) of the nuclear envelope (NE) and connects components of the cytoskeleton, including microtubules, with the inside of the nucleus such that cytoskeletal forces are transferred to the nuclear content (reviewed in Starr and Fridolfsson, 2010; Wilson and Foisner, 2010; Chang et al., 2015). The key components of the mammalian LINC complex are the trans-

membrane SUN-domain proteins, SUN1 and SUN2, which span the INM and interact with the KASH-domain nesprin proteins in the lumen of the NE. Nesprins cross the ONM and connect to cytoplasmic filaments, including microtubules. Using microtubule poisons in combination with SUN1/2 and kinesin KO MEFs, we show that the 53BP1-dependent mobility of dysfunctional telomeres is a LINC/microtubule-dependent process that promotes NHEJ. Furthermore, we document that the same 53BP1/LINC/microtubule-dependent mechanism promotes the mobility of IR-induced DSBs and contributes to their mis-repair in PARPi-treated BRCA1-deficient cells. These results establish a feature of the DDR that can lead to aberrant DNA repair when cells sustain large numbers of breaks. We argue that this potentially dangerous system is adaptive in the context of the physiological DDR, which has evolved to ensure correct DNA repair in cells with few DSBs.

RESULTS

A Standardized Method for Analysis of Dynamic Behavior of DNA Damage Foci

The mechanism of 53BP1-dependent mobility was studied using immortalized TRF2$^{F/F}$Cre-ER^{T1} MEFs expressing an mCherry-53BP1 fusion protein that contains the Tudor, UDR, and oligomerization domains of 53BP1 (Figure 1A). This mCherry fusion accumulates at DSBs and deprotected telomeres but is neither functional nor interferes with the function of the endogenous 53BP1 (Dimitrova et al., 2008).

As expected, mCherry-BP1-2 formed foci at the dysfunctional telomeres generated by Cre-mediated deletion of TRF2 (Figures 1A–1C), allowing detection of the dynamic behavior of mCherry-marked dysfunctional telomeres using 3D time-lapse microscopy and automated tracing in deconvolved images (Figure 1D; Movie S1A). Since MEF nuclei are flat (2–4 μm in the z direction compared to 15–20 μm in x and y), the data were analyzed in 2D-maximum intensity projected images.

Although the resulting traces can be corrected for the nuclear translocation and rotation (Dimitrova et al., 2008), large-scale nuclear deformation, such as expansion, contraction, folding, and twisting, also confounds the analysis. We therefore developed a standardized method to select nuclei that do not display overt distortions. The method is based on three parameters (Figure S1) applied to the data after correcting for the translocation and rotation of the nuclei as described previously (Dimitrova et al., 2008). First, because extensive distortion of a nucleus will usually shift the geometrical center (Figure S1A, type I; Movie S2A), the maximal movement of the geometrical center (MMGC) of the nucleus was evaluated (Figure S1B). Second, to identify nuclei undergoing expansion or contractions (Figure S1A, type II; Movie S2B), the maximal difference between the average distances of the foci from the geometrical center (MΔAD) was determined (Figure S1C). Third, we identified nuclei with groups of foci moving in the same direction, which could indicate nuclear folding, twisting, or rotation (Figure S1A, type III; Movie S2C). For this determination, the percentage of foci moving in the four different quadrants of the XY projections (upper right [UR]; lower right [LR]; upper left [UL]; lower left [LL]) was determined (Figures S1D and S1E).
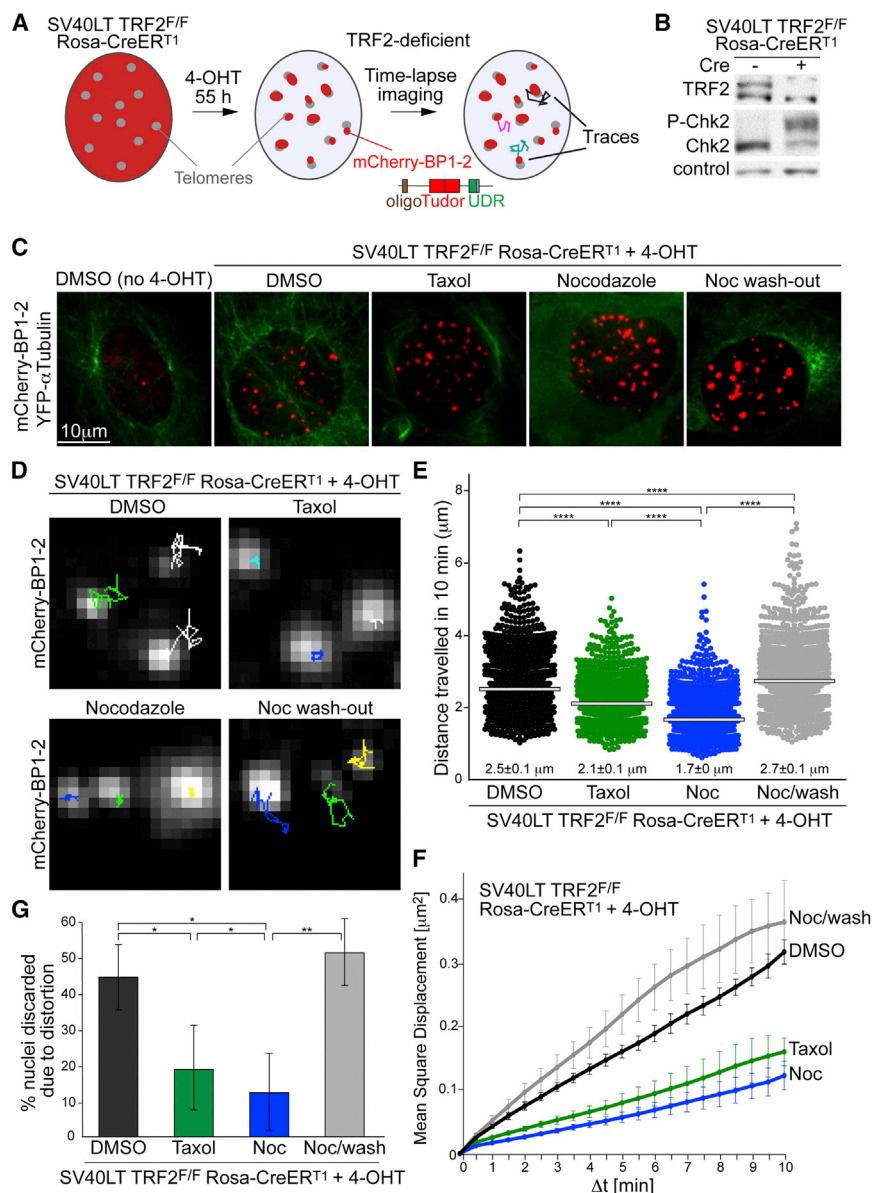
measured for dysfunctional telomeres to MSD = A + Γt$^{\alpha}$ showed an anomalous diffusion coefficient (α) of close to 1.0 (Table S1), indicating diffusive motion. The calculated diffusion coefficient (3.7 × 10^{-3} μm^2/s; Table S1) is in the range observed by others for dysfunctional mammalian telomeres (Chen et al., 2013; Cho et al., 2014), DNA damage lesions formed after UV and IR irradiation of mammalian cells (Kruhlak et al., 2006; Falk et al., 2007; Mahen et al., 2013; Becker et al., 2014), and a locus next to an I-SceI induced DSB in yeast (Miné-Hattab and Rothstein, 2012; Dion et al., 2012).

53BP1-Dependent Mobility Requires Dynamic Microtubules

We previously showed that the movement of dysfunctional telomeres is not affected by the actin drug, latrunculin A (Dimitrova et al., 2008). In contrast, when cells were incubated with the microtubule poisons Taxol or nocodazole, which stabilize and depolymerize microtubules, respectively (Figure 1C), there was a striking reduction in the mobility of the dysfunctional telomeres and the distance traveled by the telomeres was significantly smaller (Figures 1D–1F; Table S1; Movies S1A–S1C). The effect of nocodazole was completely reversed within 1 hr of its removal from the media, showing that the lack of dynamic behavior was

Sets of foci that show concerted movement will over-populate one of these quadrants, allowing detection of nuclei with distortions. Similarly, over-population of half the space in the projections (lateral, vertical, diagonal) was used to detect nuclear rotation. Using arbitrarily set thresholds for these parameters (Figure S1; see Experimental Procedures), nuclei were discarded from the analysis. In most experiments, approximately half the nuclei passed these selection criteria and were deemed to retain their shape.

Analysis of the selected nuclei showed that dysfunctional telomeres traveled a median cumulative distance of 2.5 μm in 10 min (Figures 1D and 1E; Table S1; Movie S1A), which is consistent with previous data (Dimitrova et al., 2008). The mean square displacement (MSD) increased over time, with a final MSD of 0.3 μm^2 after 10 min (Figure 1F; Table S1). Fitting of the MSD

Figure 2. SUN1 and SUN2 Promote Mobility of Dysfunctional Telomeres

(A) Immunoblots for TRF2, SUN1, SUN2, 53BP1, and phosphorylated Chk2 in the indicated MEFs at 72 hr after Hit&Run Cre.

(B) Telomere dysfunction-induced foci (TIF) assay on the MEFs described in (A). Telomeres were detected by FISH with FITC-(CCCTAA)$_3$ probe (green). Phosphorylated H2AX (top panel), 53BP1 (middle panel), and Rif1 (bottom panel) were detected by IF (red). DAPI, DNA (blue).

(C) Quantification of TIF response after Cre as assayed in (B). Cells with greater than nine TIFs were scored. Values are means and SDs of three independent experiments. p values were from an unpaired t test (see legend to Figure 1).

(D) Examples of traces of mCherry-53BP1-2 foci at 66–72 hr after Cre (see Movies S3A–S3C).

(E and F) Distribution of the cumulative distance traveled and MSDs with SDs of mCherry-BP1-2 foci in the analyzed MEFs (as in D) in four experiments, as described in Figure 1.

See also Figure S2 and Table S1.

53BP1 was detected by chromatin immunoprecipitation (ChIP) at dysfunctional telomeres in SUN1/2 DKO cells (Figures 2A–2C, S2A, and S2B). Nonetheless, the SUN1/2-deficient cells showed a significant reduction in the mobility of the dysfunctional telomeres (Figures 2D–2F and S2C; Table S1; Movies S3A and S3B). The effect of removal of SUN1 and SUN2 was at least as strong as the effect of absence of 53BP1 monitored in parallel experiments (Figures 2D–2F; Table S1; Movies S3C and S3B).

The percentage of nuclei that were discarded due to deformation was reduced in the absence of SUN1 and SUN2 (Figure S2C), implicating the LINC complex in the microtubule-mediated changes in nuclear shape. In contrast, 53BP1 had no effect on nuclear deformation (Figure S2C).

Importantly, in the TRF2 SUN1/2 TKO cells, the diminished mobility of the dysfunctional telomeres was accompanied by a reduction in their fusion (Figures 3A and 3B). Metaphase spreads of cells lacking SUN1 and SUN2 showed a 2-fold decrease in the NHEJ of dysfunctional telomeres at 84 hr (Figures 3B and S2B). The reduction in telomere fusions was also apparent from the diminished appearance of fused telomeric restriction fragments (Figures S2D and S2E). The difference in telomere fusion frequency with and without the SUN proteins was negligible when the assay was saturated at a later time point (108 hr). In contrast, in 53BP1-deficient cells telomere fusions remained infrequent at later time points, consistent with 53BP1 promoting telomere fusions through inhibition of resection as well as SUN1/2-dependent mobility. Since the absence of SUN1 alone affected telomere fusions less than absence of both SUN1 and SUN2
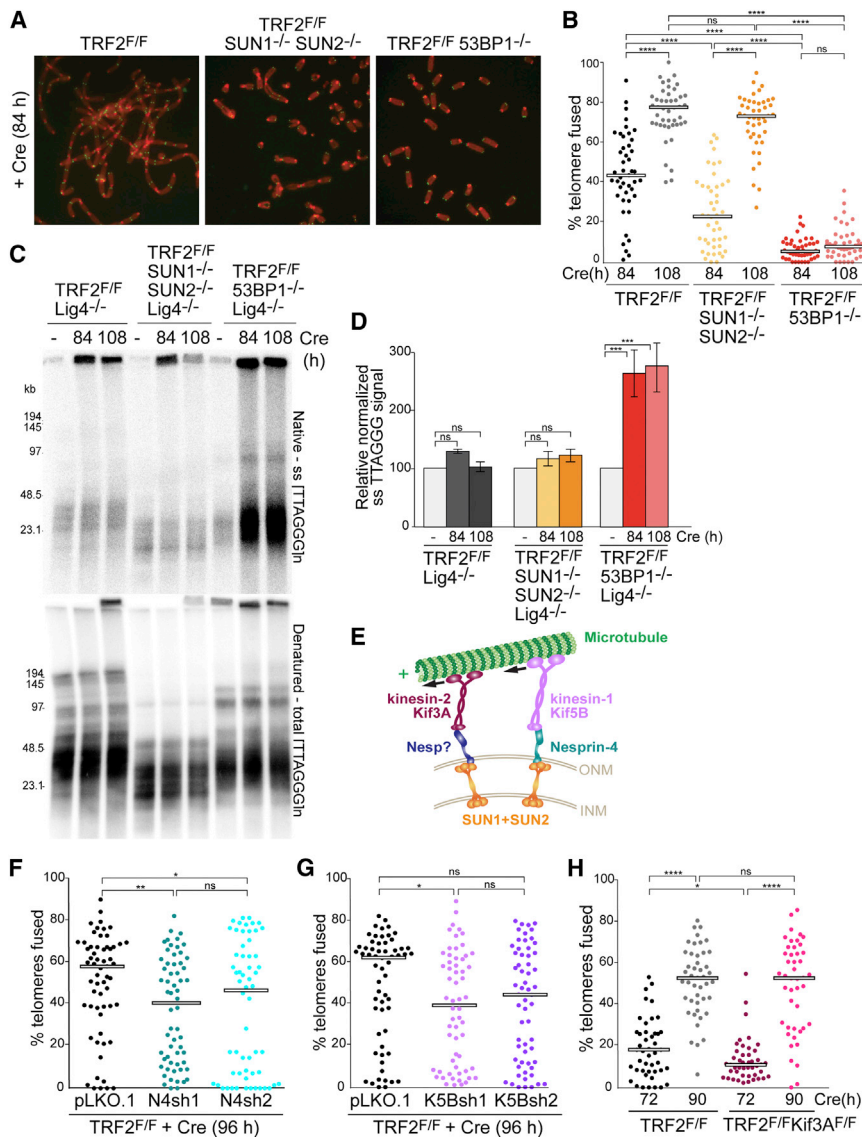
not due to a permanent toxic effect of the drug (Figures 1D–1F; Table S1; Movie S1D). Both microtubule poisons also affected the extent to which the nuclei were distorted (Figure 1G; Table S1), indicating that much of the nuclear deformation observed in these fibroblasts is microtubule-dependent.

SUN1 and SUN2 Promote the Mobility and NHEJ of Dysfunctional Telomeres

Since the involvement of microtubule dynamics suggested a link between the dysfunctional telomeres and the cytoplasm, we tested the role of the LINC complex in the movement of dysfunctional telomeres. To this end, we used SUN1 and SUN2 KO mice (Ding et al., 2007; Lei et al., 2009) to generate immortalized conditional TRF2$^{F/F}$ SUN1$^{-/-}$SUN2$^{-/-}$ MEFs. The absence of the two SUN proteins did not interfere with Chk2 phosphorylation or the formation of telomere dysfunction-induced foci (TIFs) containing γH2AX, 53BP1, and Rif1 after deletion of TRF2 and

Figure 3. The LINC Complex Promotes NHEJ of Dysfunctional Telomeres

(A) Metaphases showing telomere fusions in the indicated MEFs at 84 hr after Hit&Run Cre. Telomeres were detected by FISH with a FITC-(CCCTAA)$_3$ probe (green). DNA, DAPI (red).

(B) Distribution of telomere fusions as in (A) at 84 and 108 hr after Cre. Dots represent % fusions in individual metaphases. Bars represent the median of telomere fusions in 15 metaphases for three independent experiments (45 metaphases). p values from unpaired t test (see legend to Figure 1).

(C) In-gel assay for single-stranded telomeric DNA. Telomeric overhangs detected in situ with end-labeled ^{32}P-(AACCCT)$_4$ in MboI-digested genomic DNA from the indicated MEFs at 84 and 108 hr after TRF2 deletion (top panel). Bottom: the DNA was denatured in situ and rehybridized with the same probe to determine the total telomere DNA.

(D) Quantification of relative overhang signal as detected in (C). Values represent means for four independent experiments with SDs. The ss telomeric signal was normalized to the total telomeric DNA in the same lane. For each MEF line, the normalized no Cre value of cells was set at 100 and the post-Cre values are given relative to this value. Two-way ANOVA for multiple comparisons were used to perform statistical analysis. For p value symbols see legend to Figure 1.

(E) Schematic of the LINC complex and microtubules.

(F and G) Quantification of telomere fusions in TRF2$^{F/F}$ MEFs treated with shRNAs to nesprin-4 or Kif5B 96 hr after Cre and analyzed as in (A) and (B). Bars represent the median % of telomeres fused in three independent experiments (20 metaphases each).

(H) Quantification of telomere fusions in TRF2$^{F/F}$ RsCre-ER^{T1} and TRF2$^{F/F}$ Kif3A$^{F/F}$ RsCre-ER^{T1} MEFs 72 and 90 hr after 4-OHT, as in (A) and (B). See also Figures S2 and S3.

(Figure S2F), we conclude that the SUN proteins have partially overlapping functions in this pathway.

We verified that the deficiency in telomere fusion in the SUN1/2 KO was not due to increased resection using a quantitative assay for the amount of ssTTAGGG repeats after deletion of TRF2 from SUN1/2-deficient cells. In Lig4$^{-/-}$ MEFs (TRF2/SUN1/SUN2/Lig4 quadruple KO), which are a good system for detection of resection because the telomeres remain free, there was no great increase in the overhang signal after TRF2 deletion (Figures 3C and 3D), indicating that resection remained repressed. Parallel deletion of TRF2 from 53BP1$^{-/-}$ Lig4$^{-/-}$ cells showed the substantial increase in overhang signal expected from the role of 53BP1/Rif1 in repression of resection (Lottersberger et al., 2013). These data, together with the normal localization of Rif1 at dysfunctional telomeres in SUN1/2 DKO cells (Figures 2B and 2C), supports the idea that SUN1 and SUN2 are dispensable for the protection of DSBs from resection and

act independently from Rif1. We propose, therefore, that SUN1 and SUN2 promote the c-NHEJ of telomeres by increasing their dynamic behavior.

Nesprin-4 and Kinesins Contribute to NHEJ of Dysfunctional Telomeres

Since the SUN proteins are connected to the cytoskeleton through nesprins (reviewed in Starr and Fridolfsson, 2010) (see Figure 3E) and SUN1/2-deficient cells lack nesprin-1, nesprin-2, nesprin-3, and nesprin-4 at the NE (Crisp et al., 2006; Padmakumar et al., 2005; Ketema et al., 2007; Lei et al., 2009; Roux et al., 2009), we tested shRNAs to nesprins for an effect on NHEJ of dysfunctional telomeres. Two shRNAs targeting nesprin-4 lowered the frequency of telomere fusions without affecting cell proliferation or the DDR (Figures 3F, S3A, and S3B).

As nesprin-4 is known to interact with the plus-end directed microtubule motor kinesin-1 (Figure 3E), we tested shRNAs to

the Kif5B subunit of kinesin-1 for an effect on the NHEJ of dysfunctional telomeres. Two shRNAs to Kif5B lowered the frequency of telomere fusions at an early time point without affecting the proliferation or the DDR upon telomere deprotection (Figures 3G, S3C, and S3D). In addition, two shRNAs to the kinesin-2 subunit Kif3A resulted in a reduced frequency of telomere fusions (Figures S4E–S4G). Since kinesin-2 had not previously been shown to cooperate with nesprin-4, we generated TRF2$^{F/F}$Kif3A$^{F/F}$ MEFs to further verify the shRNA data. Consistent with the shRNA results, MEFs lacking Kif3A showed a significant reduction in the efficiency of telomere fusions after TRF2 deletion (Figures 3G, S3H, and S3I). These data suggest that 53BP1-mediated mobility of dysfunctional telomeres likely involves redundant action by the microtubule motors kinesin-1 and kinesin-2, as well as nesprin-4 and possibly other nesprins.

Phosphorylation Sites in 53BP1 Required for Telomere Mobility

As a version of 53BP1 lacking its N-terminal S/TQ sites (53BP1-28A) fails to induce chromatin mobility (Lottersberger et al., 2013), we determined which S/TQ sites are involved in this process. We generated a collection of S or T to A mutations at the S/TQ positions in a C-terminally truncated version of 53BP1 that lacks the BRCT domain (53BP1DB; Figure 4A) (Bothmer et al., 2011) and behaves like wild-type 53BP1 in the context studied here (Lottersberger et al., 2013; Zimmermann et al., 2013). Through the analysis of the mutants, we identified one mutant, referred to as 53BP1ΔMOB, which appeared to be a separation-of-function mutant specifically deficient in the ability of 53BP1 to promote mobility but proficient in blocking resection (Figure 4A). 53BP1ΔMOB recruited Rif1 to sites of DNA damage and was able to interact with PTIP, which was expected since the region of mutated S/TQ sites falls outside the previously mapped Rif1 and PTIP interacting regions (Figures 4A and S4A–S4C) (Munoz et al., 2007; Escribano-Díaz et al., 2013). Consistent with its binding to Rif1, the 53BP1ΔMOB mutant was proficient in repressing hyper-resection at telomeres after TRF2 deletion in TRF2$^{F/F}$ 53BP1$^{-/-}$ Lig4$^{-/-}$ cells (Figures S4D and S4E).

Despite the normal interactions with Rif1 and PTIP, the ability of 53BP1ΔMOB to promote telomere fusions upon complementation of 53BP1 deficiency was significantly reduced (Figure 4B). However, 53BP1ΔMOB promoted telomere fusion similar to 53BP1DB in SUN1$^{-/-}$ SUN2$^{-/-}$ 53BP1$^{-/-}$ cells (Figures 4B and S4F), suggesting that the 53BP1ΔMOB is only deficient in a function that requires SUN1/2. Time-lapse imaging showed that 53BP1ΔMOB is completely defective in promoting the increased mobility of dysfunctional telomeres resulting in dynamics that are indistinguishable from cells transduced with the empty vector or the 53BP1Δ28A mutant (Figures 4C and S4G; Table S1). In contrast, 53BP1ΔPTIP showed no defect in promoting mobility of dysfunctional telomeres (Figures 4C and S4G; Table S1). Thus, the ability of 53BP1 to promote mobility of dysfunctional telomeres likely involves an interaction that depends on phosphorylation of one or more of the ST/Q sites in the MOB domain. The identity of the MOB domain interacting partner is unknown. It is not excluded that this domain interacts with SUN1 and SUN2 but this interaction was not detected by

mass spectrometry (Di Virgilio et al., 2013) and ChIP failed to reveal SUN1 and SUN2 at dysfunctional telomeres (Figure S2A).

PTIP Is Not Required for 53BP1-Dependent Mobility

To determine whether PTIP contributes to the 53BP1-dependent mobility, TRF2 and PTIP co-deletion in SV40LT immortalized TRF2$^{F/F}$ PTIP$^{F/F}$ MEFs was analyzed. Absence of PTIP did not affect cell proliferation or the DDR at the dysfunctional telomeres (Figures S5A–S5D). In the PTIP-deficient setting, the distances traveled and MSD of the dysfunctional telomeres was equal to that of PTIP containing control cells (Figures 4D, 4F, and S5E; Table S1; Movies S4A–S4C). Moreover, the analysis of the telomere overhangs showed that PTIP deficiency did not affect the resection at dysfunctional telomeres (Figures S5F and S5G), supporting the previous conclusion that 53BP1-dependent protection from resection is primarily dependent on Rif1 (Zimmermann and de Lange, 2014). Nonetheless, as previously shown (Callen et al., 2013), telomere fusions appeared slightly delayed when PTIP was deleted (Figures 4G, S5F, and S5G). Consistent with these results, the 53BP1ΔPTIP mutant displayed a mild defect in promoting telomere fusions but appeared unaffected with regard to protection from resection and the induction of mobility (Figures 4B, 4C, S4D, and S4E).

53BP1/LINC/Microtubule-Dependent Mobility of IR-Induced DSBs

Despite their resemblance to DSBs, dysfunctional telomeres could be argued to be different from chromosome-internal DNA breaks. We therefore tested whether genome-wide DSBs are subject to the 53BP1/LINC/microtubule-dependent changes in dynamics. To this end, we analyzed the mobility of the mCherry-BP1-2 foci after induction of ~100 DSBs with 2.75 Gy IR (Rothkamm and Löbrich, 2003) in wild-type, SUN1$^{-/-}$SUN2$^{-/-}$, and 53BP1$^{-/-}$ MEFs. As expected, the Chk2 phosphorylation and formation of γ-H2AX foci were not affected by the genotype of the cells (Figures 5A–5C). The IR-induced mCherry-53BP1-2 foci showed a cumulative distance traveled and an MSD comparable to the MSD of dysfunctional telomeres. This dynamic behavior was strongly diminished in absence of 53BP1 or the SUN proteins and upon treatment with Taxol (Figures 5D–5G; Table S1; Movies S5A–S5D). Therefore, we conclude that the 53BP1/LINC/microtubules pathway promotes the mobility of chromosome-internal DSBs as it does at dysfunctional telomeres.

Undamaged Chromatin Is Minimally Affected by DSBs

We next asked whether the presence of mobile DSBs changes the dynamics of the global chromatin. To address this question, we monitored the mobility of fully functional telomeres, marked with eGFP-TRF1 in cells with and without IR-induced DSBs. The IR was delivered at 2.75 Gy, which induces ~1 DSB/60 Mb (~100 DSBs per cell, see above). Since the 80 telomeres of the mouse genome represent ~4 Mb (~0.1% of the genome), telomeres are not expected to contain DSBs after 2.75 Gy. Nonetheless, the eGFP-marked telomeres showed a very slight but statistically significant increase in the cumulative distance traveled (Figure S6). Moreover, their MSD and diffusion coefficient were slightly increased, although much less than when
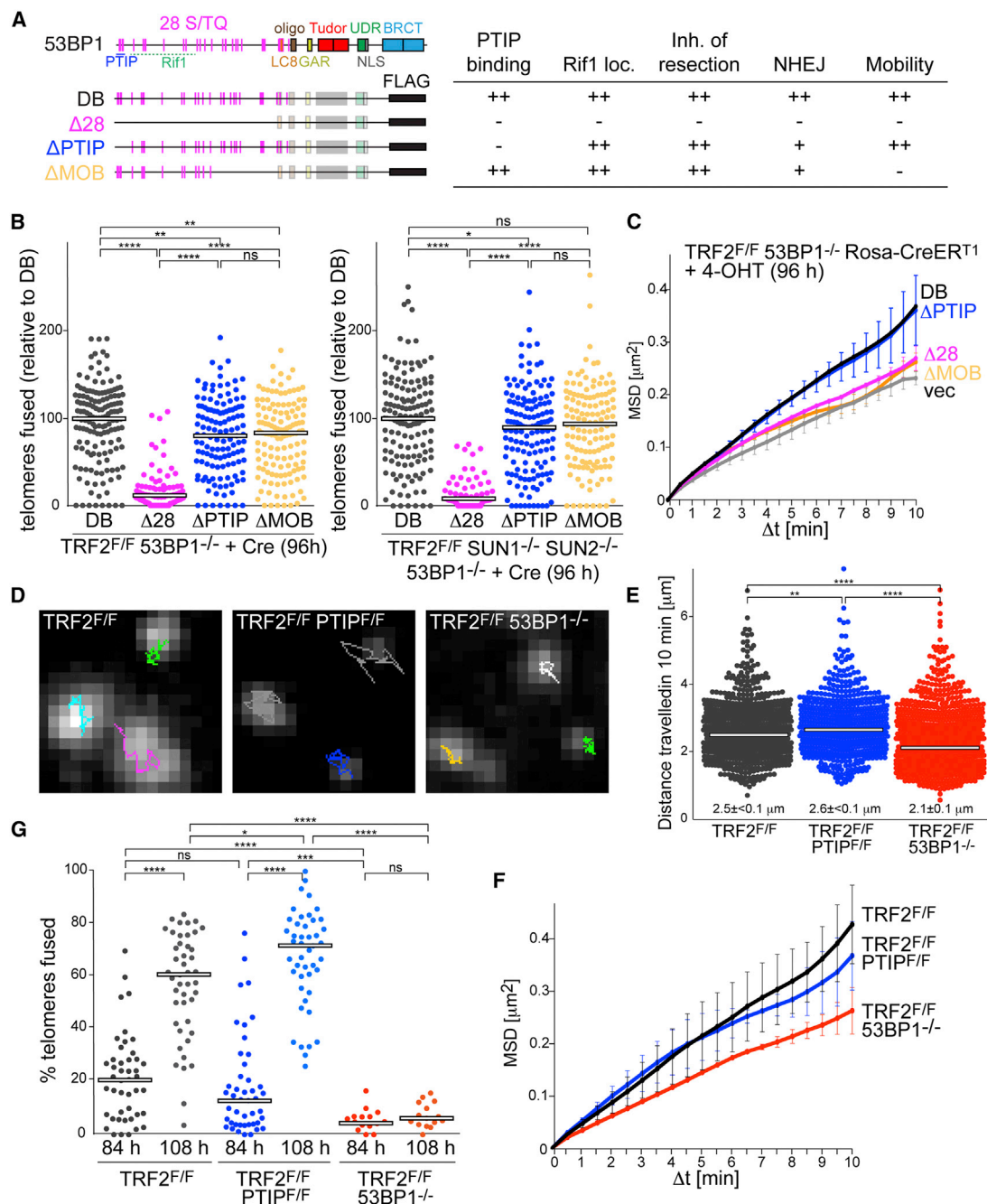
Figure 4. The Mobility Domain of 53BP1, but Not PTIP, Is Required for Mobility of Dysfunctional Telomeres

(A) Schematic of 53BP1, S/TQ site mutations, and their phenotypes.

(B) Quantification of telomere fusions in the indicated MEFs complemented with the indicated 53BP1 alleles 96 hr after TRF2 deletion with Hit&Run Cre (as in Figure 3). Data from >70 metaphases analyzed in four independent experiments. For each experiment, the median fusion frequency for 53BP1DB was set to 100 and all other values were normalized to this frequency.

(C) MSDs with SDs of mCherry-BP1-2 foci detected in the TRF2-deleted $53BP1^{-/-}$ RsCre-ER^{T1} MEFs expressing the indicated 53BP1 alleles. Data from three independent experiments.

(D) Examples of traces of mCherry-53BP1-2 foci at 66–72 hr after Cre in the indicated MEFs (see Movies S4A–S4C).

(E and F) Distribution of the cumulative distance traveled and MSDs with SEMs of mCherry-BP1-2 foci in the indicated MEFs (as in Figure 1). Bars represent medians of the cumulative distance traveled by >500 foci in two experiments and numbers indicate the averages and SEMs of the two median values obtained in two independent experiments.

(G) Quantification of telomere fusions in the indicated MEFs at 84 and 108 hr after Cre (as in Figure 3).

See also Figures S4 and S5 and Table S1.
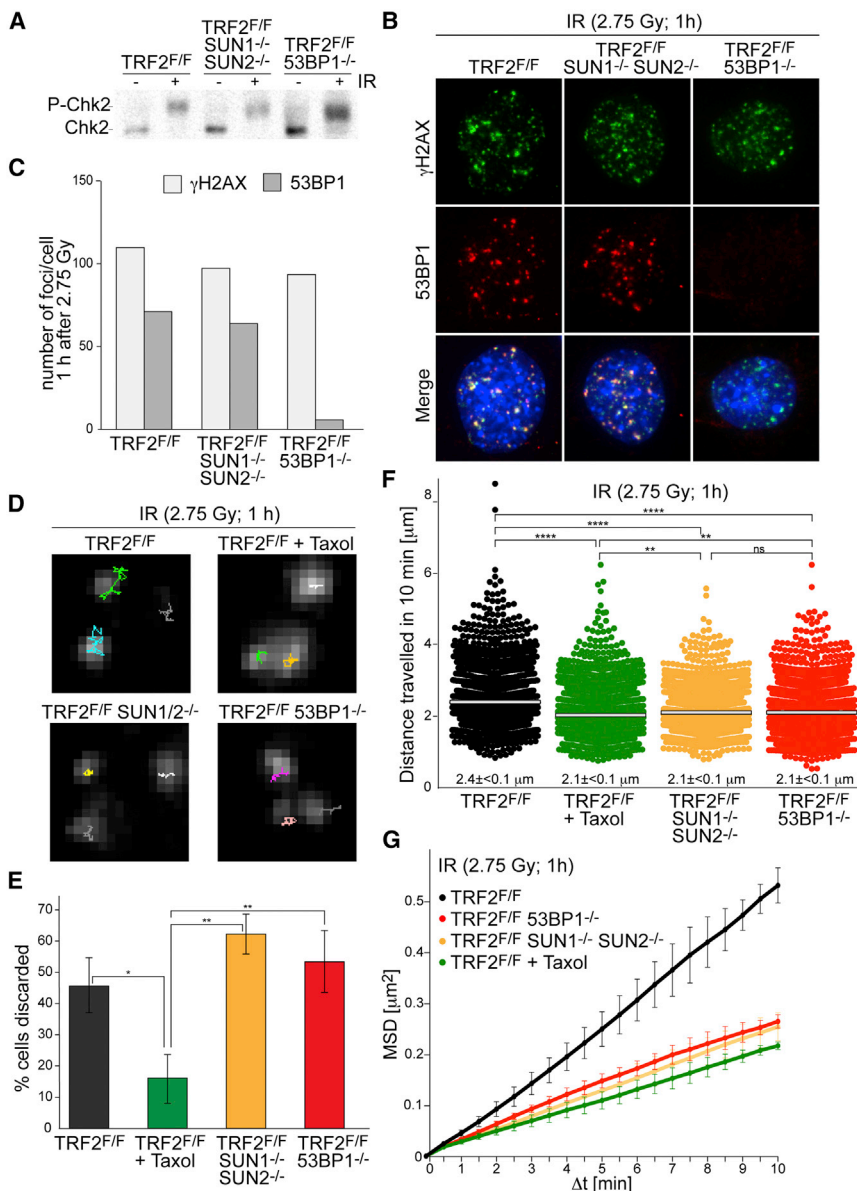
Figure 5. 53BP1/LINC/Microtubule-Promoted Mobility of IR-Induced DSBs

(A) Immunoblot for phosphorylation of Chk2 (as in Figure 2A) in the indicated MEFs at 1 hr after 2.75 Gy IR.

(B) IF for γH2AX (green) and 53BP1 (red) for cells treated as in (A). DAPI, DNA (blue).

(C) Quantification of IR-induced γ-H2AX and 53BP1 foci as assayed in (B).

(D) Examples of 10 min traces of mCherry-53BP1-2 foci at 1 hr after IR of the cells described in (A) with or without 20 μM Taxol (see Movies S5A–S5D).

(E–G) Percentage of cells discarded, distribution of the cumulative distance traveled, and MSDs with SDs of mCherry-BP1-2 foci detected as (D) and (E) (as in Figure 1). Data from three independent experiments.

See also Figure S6 and Table S1.

DNA ends that are at a distance. One setting in which this process may be relevant is the formation of radial chromosomes in PARPi-treated BRCA1-deficient cells. Radial formation involves the joining of a DNA end from one chromosome with a break in another chromosome, which may be at a distance and therefore would require spatial exploration for joining. We therefore tested whether the 53BP1-dependent mobility contributes to the mis-rejoining when many S phase DSBs are induced with PARPi and HDR is impaired. Experiments with cells containing fluorescently labeled geminin to reveal their cell-cycle stage showed that IR-induced DSBs become mobile in S/G2 as well as in G1 (Figures S7A–S7D).

As previously shown, when BRCA1 shRNA-treated cells were incubated with the PARP inhibitor olaparib, a significant number of mis-rejoined chromosomes was formed and this phenotype

the telomeres were dysfunctional (Figure S6; Table S1; Movies S6A–S6D). These results indicate that while the chromatin dynamics primarily affects sites of DNA damage, there is also a minor increase in the mobility of undamaged chromatin, consistent with a previous report (Zidovska et al., 2013).

When the eGFP-TRF1 marker was used to detect nuclear deformations, the incidence of distorted nuclei was not affected by deletion of TRF2 (Figure S6B; Table S1), indicating that microtubule dynamics distort nuclei regardless of the presence of DNA damage.

Chromatin Mobility Promotes DSB Mis-repair in BRCA1-Deficient Cells

We considered that for genome-wide DSBs, the increased mobility of the chromatin could promote the joining of unrepaired

was repressed by deletion of 53BP1 (Figures 6A–6C). Importantly, SUN1$^{-/-}$SUN2$^{-/-}$ MEFs also diminished the formation of aberrantly repaired chromosomes (Figures 6A–6C) and the mis-rejoining events were strongly reduced by Taxol (Figure 6D). The effect of Taxol was not due to diminished PARP inhibition, since PARPi/Taxol-treated cells showed no parsylation in response to H_2O_2 (Figure S7E). Taxol did not further reduce either the mobility or the chromosome mis-rejoining events in absence of SUN1 and SUN2 (Figures 6E and S7F–S7H; Table S1), supporting the view that the SUN proteins and microtubules act in the same pathway to promote chromatin mobility and aberrant DNA repair. Importantly, SUN1/2 deficiency also diminished the lethality of PARPi treatment in BRCA1-deficient cells (Figures 6F and S7I). As expected, the absence of 53BP1 rescued the lethality of PARPi treatment to a greater extent,
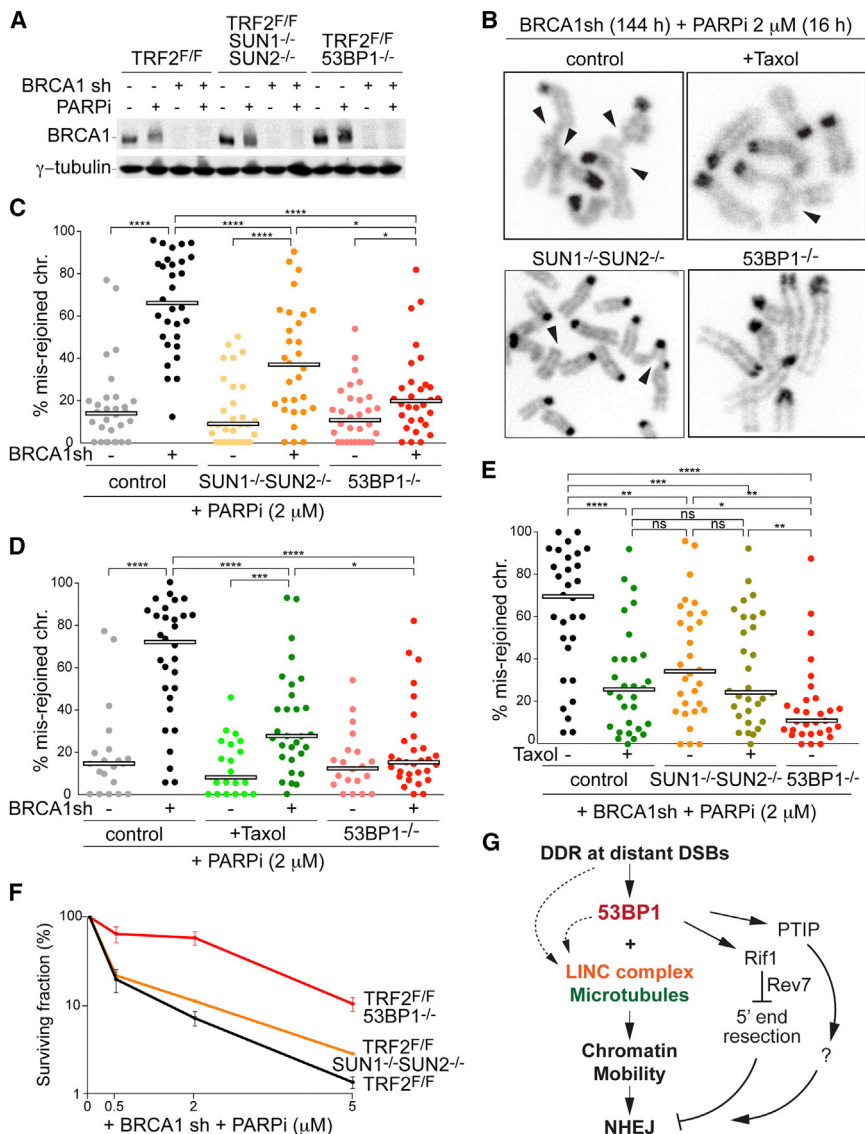
Figure 6. SUN1/2 and Dynamic Microtubules Promote Radial Formation

(A) Immunoblots for BRCA1 and γ-tubulin in the indicated MEFs (as in Figure 2A) at 144 hr after infection with BRCA1 shRNA or empty vector. Olaparib was added 16 hr before analysis.

(B) Representative mis-rejoined chromosomes (arrowheads). DNA stained with DAPI.

(C) Quantification of mis-rejoined chromosomes in the indicated MEFs (as in A), analyzed as in (B). Each dot represents a metaphase. Bars represent the median of mis-rejoined chromosomes in three independent experiments (10 metaphases each). p values as in Figures 1A and 3B.

(D) Quantification of mis-rejoined chromosomes in the indicated MEFs 18 hr with or without Taxol as in (C).

(E) Quantification of mis-rejoined chromosomes in each metaphase in the indicated MEFs with or without Taxol as described in (C) and (D). All cells used in (A)–(F) are TRF2$^{F/F}$.

(F) Quantification of colony formation in the indicated cells infected with BRCA1 shRNA and treated with or without olaparib for 7 days. The curves represent the average and SEMs of two independent experiments.

(G) Schematic of the role of 53BP1 in NHEJ of distant DSBs. In addition to controlling of DNA end processing, 53BP1 can affect NHEJ by increasing the mobility of DSBs. The mobility of DSBs is dependent on the LINC complex and microtubule dynamics. Dashed arrows indicate the possibility that the DDR affects the LINC complex and microtubules independent of 53BP1.

See also Figure S7.

NE in preparation for meiosis I, the DSB mobility recorded here is not overtly associated with clustering or NE targeting.

In the experimental settings analyzed here, the spatial exploration of DSBs promotes their pathological joining by NHEJ. DSB mobility enhanced telomere-telomere fusions forming dangerous dicentric chromosomes and similarly, it promoted the mis-repair of PARPi-induced DSBs generating lethal radial chromosomes. Given these fatal outcomes, a major question is why this pathway is allowed to act on DSBs. Below, we propose that the enhanced mobility of DSBs represents a mechanism to restore the connection between DNA ends that have lost their proper interaction. We argue that this mechanism can counteract ectopic repair when DSBs are rare, as is the case under physiological conditions. On the other hand, DSB mobility will promote mis-repair under experimental conditions when a high number of DSBs are generated at the same time.

consistent with the multiple mechanisms by which 53BP1 affects DSB repair (Figure 6G).

DISCUSSION

These results establish that DSBs show altered dynamic behavior in mammalian nuclei. The mobility and roaming of damaged chromatin requires the MOB domain in 53BP1, the SUN1/2 components of the LINC complex, and dynamic microtubules. In addition, data on telomere fusions implicated plus-end directed microtubule motors (kinesin-1 and kinesin-2) and at least one of the nesprin proteins in this process. The LINC complex contributes to the dynamic behavior of specific chromosomal loci, including telomeres, during bouquet formation in many eukaryotes (reviewed in Shibuya and Watanabe, 2014). However, the process acting on DSBs is different from bouquet formation. While the bouquet configuration bundles loci at one area of the

How DSB Mobility Could Prevent Repair Errors in G1 and S/G2

It is reasonable to assume that 53BP1 did not evolve to promote the fusion of dysfunctional telomeres and mis-repair of DSBs in
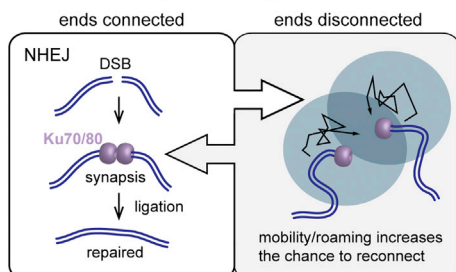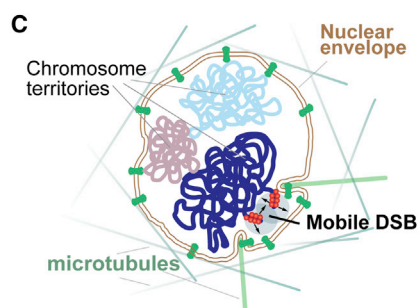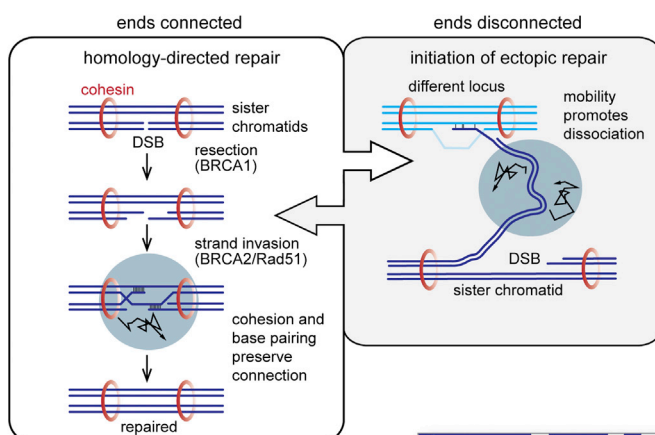
Figure 7. Proposed Function and Mechanism of 53BP1-Dependent Mobility of DSBs

(A and B) Proposed function for 53BP1-dependent mobility in promoting correct DSB repair. (A) G1: mobility of DNA ends that have lost their association could promote their rejoining, thereby promoting NHEJ. (B) S/G2: if a DNA end loses connection with the sister chromatid and invades an ectopic locus, DSB mobility could disrupt this aberrant interaction and promote correct HDR. If the DSB is being repaired correctly using HDR on the sister chromatid, mobility will not dissociate the ends because of the presence of cohesin and base-pairing.

(C) Proposed models for the mechanism of 53BP1/LINC/microtubule-dependent mobility of DSBs. The enlarged part of the nucleus shows 53BP1 (red) at a DSB with the ends separated. One end (top) portrays a model in which 53BP1 has a physical connection with the LINC complex (green). The LINC complex connects to dynamic microtubules and thereby moves the LINC-bound 53BP1-covered DNA end. The other end (bottom) portrays a model in which there is no physical connection between the LINC complex and 53BP1. The LINC complex associates with microtubules that "poke" the nucleus. The 53BP1-associated chromatin moves more readily even when not at the periphery, perhaps because 53BP1 alters the flexibility of the chromatin fiber. See text for discussion.

PARPi-treated BRCA1-deficient cells. Instead, we propose that 53BP1 has gained the ability to promote DSB mobility to facilitate correct repair (Figure 7). We imagine two settings where increased chromatin mobility at a DSB would be advantageous. The first setting is in G1 when a DSB is formed and its repair by Ku70/80-dependent c-NHEJ is the preferred mechanism to reestablish the integrity of the genome (Figure 7A). If Ku loading fails or synapsis does not occur, the DNA ends might become spatially separated. For instance, chromatin-remodeling and nucleosome eviction at DSBs (reviewed in Peterson and Almouzni, 2013) may drive the two DNA ends apart. If the separated ends are mobile, their increased spatial exploration could reconnect them and promote their joining.

The second setting in which mobility of damaged chromatin could prevent repair errors is after DNA replication (Figure 7B).

In S/G2, DSBs can be repaired by HDR using the sister chromatid as the template. However, if the DNA topology is unfavorable, one DNA end (or both) could lose its attachment to the sister chromatid and initiate ectopic repair on a different locus (Figure 7B). Mobility of the chromatin near the DSB could help to disconnect the wandering DNA end from an ectopic locus where it is not held down by cohesin and where base-pairing will be limited. In contrast, chromatin mobility of DSBs is less likely to interrupt HDR on the sister because of the stabilizing effects of cohesion and base-pairing.

The proposed role of DSB mobility in counteracting ectopic interactions is analogous to what has been proposed for the mobility of the chromosome pairing centers in *Caenorhabditis elegans* meiosis (Sato et al., 2009). Sato et al. (2009) argued that this process preferentially disrupt pairing of non-homologous chromosomes since paired homologs will have a greater ability to resist forces. Although the system described here is different from the meiotic events, both regulatory pathways may have evolved to provide a mechanism aimed to distinguish weak non-homologous interactions from the stronger connection afforded by homology.

A key consideration with regard to the role of 53BP1 in DSB repair is that the mammalian DDR did not evolve to handle

hundreds of DSBs occurring at the same time. In vivo, the majority of cells in primate brain and liver show no evidence of DSBs and only 10% of the cells have one or two 53BP1 foci (Fumagalli et al., 2012); U. Herbig, personal communication), indicating that the occurrence of multiple DSBs in one nucleus is rare in post-mitotic tissues. Furthermore, in MEFs that are in S phase, where DSBs are expected to be more frequent, <20% of the nuclei have five or more 53BP1 foci and none showed more than ten (Wu et al., 2010). This number of potential S phase DSBs may be an overestimate because 53BP1 foci can form at a variety of DNA lesions. These observations argue that the 53BP1-mediated mobility of DSBs is unlikely to cause chromosomal aberrations unless cells experience an exogenous genotoxic insult.

Models for the Mechanism by Which DSB Mobility Is Generated

We are considering two general types of models for how 53BP1, the LINC complex, and microtubules promote mobility (Figure 7C). In the first model, there is a physical connection between the 53BP1-marked chromatin and a LINC complex that interacts with microtubules. In the second model, no such connection exists.

Although we have not been able to establish a physical interaction between 53BP1 and the SUN proteins, it is not excluded that 53BP1 directs DSBs to the LINC complex. If 53BP1 interacts with the LINC complex, kinesin- and microtubule-dependent mobility of the LINC complex could alter the dynamic behavior of DSBs. The lack of clear peripheral localization of DSBs is not a strong argument against this model since the nuclei we have studied are flat, positioning most of the chromatin fairly close to the NE. Furthermore, NE invaginations could allow a connection of a non-peripheral DSB with the LINC complex. We note that the recorded trajectories and the diffusive behavior of DSBs gleaned from the MSD curves argue against the direct interaction model. However, if the engagement is short-lived and takes place in iterative rapid steps, the outcome may resemble diffusive behavior rather directed movement.

Nonetheless, we favor a second type of model in which no physical connection occurs between 53BP1 and the LINC complex. In this model, the role of the LINC complex is to transduce microtubule forces onto the chromatin in an untargeted manner. This process may be analogous to the microtubule-mediated fenestration of the nuclear envelope in prophase, which is in part mediated by the SUN proteins (Turgay et al., 2014). Random "poking" of the nucleus in response to DNA damage would explain why the global chromatin becomes slightly more dynamic in cells with DSBs but how this process is activated by the DNA damage response remains to be determined. It is also unclear whether the visco-elastic properties of chromatin and the resistance of the lamin network allow force propagation over the required distance.

How could microtubule forces specifically increase the mobility of DNA damaged loci in absence of a connection between 53BP1 and the LINC complex? The simplest explanation would be that 53BP1, through a factor that binds to the MOB domain, changes the flexibility of the chromatin fiber containing the DSB. Increased flexibility of the large chromatin domain containing 53BP1 could render it more sensitive to the microtubule forces transduced through the NE. Indeed, chromatin that contains DSBs shows a decreased density as determined by EM and appears to expand (Kruhlak et al., 2006), attributes that could be consistent with a change in the flexibility of the chromatin fibers.

Implications

This study revealed that mammalian cells use microtubules in the cytoplasm to promote the mobility of sites of DNA damage in the nucleus. Although some of the molecular details of this process remain to be determined, the main players, including the MOB domain of 53BP1, the LINC complex, kinesins, and microtubules are now known, allowing further investigation. The results show that in cells with many DSBs, the induced mobility of the damaged chromatin can promote aberrant DSB repair events, including the fusion of dysfunctional telomeres and formation of radial chromosomes in PARPi-treated BRCA1-deficient cells. Two main issues warrant attention in the near future. First, one prediction from our findings is that curbing microtubule dynamics with taxanes might limit the efficacy of PARPi-treatment of HR-deficient cancers. Thus, when a combination of taxanes with olaparib or other DNA-damaging agents (e.g., platin drugs) is being considered, the effect of taxanes on the efficacy of genotoxic drugs merits further testing. Second, it will be of interest to test our proposal that the 53BP1-dependent mobility of DSBs can prevent DNA repair errors under normal physiological settings when DSBs are rare.

EXPERIMENTAL PROCEDURES

Live-Cell Imaging and Identification of Distorted Nuclei

Dysfunctional telomeres were visualized using mCherry-BP1-2 as described previously (Dimitrova et al., 2008). Images were deconvolved and 2D-maximum intensity projection images were obtained using SoftWoRx software. Tracking of mCherry-BP1-2 foci was performed with ImageJ software on at least ten cells per condition. Cells were registered by the StackReg plugin using Rigid Body (Thévenaz et al., 1998) and particles were tracked using the Mosaic Particle Detector and Tracker plugin (Sbalzarini and Koumoutsakos, 2005) with the following parameters for particle detection and tracking: radius = 1–2 pixels; cutoff = 1–2 pixels; percentile = 6; link range = 1; displacement = 5 pixels. The x and y coordinates of each trajectory were used for further calculation. All mCherry-BP1-2 foci in a cell that were continuously tracked for at least 19 out of 20 frames were analyzed. The analysis of the eGFP-TRF1-marked telomeres was similarly conducted using the following parameters: radius = 1 pixel; cutoff = 1 pixel; percentile = 8–12; link range = 1; displacement = 5 pixels.

The average x and y values of all the foci was calculated in each frame as the geometrical center (GC) and normalized over the GC$_{t=0}$. The distance traveled by the GC between each time points t = b and t = a was calculated as movement of geometrical center

$$MGC_{b-a} = \sqrt{\left(x^{GC}_{t=b} - x^{GC}_{t=a}\right)^2 + \left(y^{GC}_{t=b} - y^{GC}_{t=a}\right)^2},$$

and the maximal MGC (MMGC) for each cell was identified. Cells were discarded if their MMGC exceeded the arbitrary threshold of 2, or if their MMGC exceeded the secondary threshold of 1 and another parameter was also above threshold.

The difference of the average distances of all the i foci in the cell and $GC_{t=0}$ (ΔAD) between each time points $t = b$ and $t = a$ was calculated as

$$\Delta AD_{b-a} = \left| \left(\frac{\sum_{i=1}^{n} \sqrt{\left(x_{t=b}^{i} - x_{t=0}^{GC}\right)^2 + \left(y_{t=b}^{i} - y_{t=0}^{GC}\right)^2}}{n} \right) - \left(\frac{\sum_{i=1}^{n} \sqrt{\left(x_{t=a}^{i} - x_{t=0}^{GC}\right)^2 + \left(y_{t=a}^{i} - y_{t=0}^{GC}\right)^2}}{n} \right) \right|,$$

and the maximal ΔAD ($M\Delta AD$) for each cell was identified. Cells were discarded if $M\Delta AD$ exceeded the arbitrary threshold of 2, or if $M\Delta AD$ exceeded the secondary threshold of 1 and another parameter was also above threshold.

Finally, the trajectories traveled by each focus i per cell, relatively to the GC, were normalized to the coordinates $x_{t=0}^{i}$ and $y_{t=0}^{i}$ and projected together on a XY plane. The percentage of foci in each quadrant was calculated for each time frame: upper right (UR(%)), lower right (LR(%)), upper left (UL(%)), lower left (LL(%)) and the average of these values during the time lapse was derived. Laterality (LAT (%)), verticality (VER (%)), and diagonality (DIA (%)) were calculated for each time frame as:

$$LAT(\%) = \left| \left(\left((UR(\%) + LR(\%)) \div 100 \right) - 0.5 \right) \div 0.5 \right| \times 100,$$

$$VER(\%) = \left| \left(\left((UR(\%) + UL(\%)) \div 100 \right) - 0.5 \right) \div 0.5 \right| \times 100,$$

$$DIA(\%) = \left| \left(\left((UR(\%) + LL(\%)) \div 100 \right) - 0.5 \right) \div 0.5 \right| \times 100,$$

and the average of these values during the time lapse were derived. Cells were discarded if UR, LR, UL, LL, LAT, VER, or DIA exceeded the arbitrary threshold of 40%, or if they exceeded the secondary threshold of 30% and another parameter was also above threshold.

the Cumulative Distance traveled in 10 min by each of the foci i (CD^i) was calculated relative to the GC, as previously described (Dimitrova et al., 2008), as

$$CD^i = \sum_{t=1}^{20} \sqrt{\left((x_i^t - x_t^{GC}) - (x_{t-1}^i - x_{t-1}^{GC}) \right)^2 + \left((y_t^i - y_t^{GC}) - (y_{t-1}^i - y_{t-1}^{GC}) \right)^2}.$$

Mean square displacement (MSD) was calculated as

$$MSD(\Delta t) = \frac{1}{n} \times \sum_{i=1}^{n} D_i(\Delta t)^2,$$

where

$$D_i(\Delta t) = \sqrt{\left((x_t^i - x_t^{GC}) - (x_{t-\Delta t}^i - x_{t-\Delta t}^{GC}) \right)^2 + \left((y_t^i - y_t^{GC}) - (y_{t-\Delta t}^i - y_{t-\Delta t}^{GC}) \right)^2}.$$

All data output in pixels (standard ImageJ output) were converted to meters by the formula, 1 pixel = 0.215 μm, based on the characteristics of the objective.

Diffusion coefficient D was calculated as

$$D = m/4,$$

where m is the slope of the MSD after fitting to a linear curve. The anomalous diffusion coefficient α was derived using MATLAB by the fitting of MSD to the diffusion model function:

$$MSD = A + \Gamma t^{\alpha}.$$

For cumulative distance, statistical analysis was performed using Prism Software applying the Mann-Whitney test.

Other Experimental Procedures

All procedures for derivation of MEFs, cell treatments, plasmids, shRNAs, immunoblotting, IF, IF-FISH, analysis of metaphase chromosomes, in-gel analysis of telomeric DNA, co-immunoprecipitation, ChIP, and mutagenesis were performed using previously published standard procedures. The mutated 53BP1 alleles were as follows: 53BP1ΔPTIP (S6A, S13A, S25A, S29A) and 53BP1ΔMOB (S674A, T696A, S698A, S784A, S831A, T855A, S892A, S1068A, S1086A, S1104A, T1148A, S1171A, S1219A).

Detailed experimental procedures are given in the Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, one table, and six movies and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.09.057.
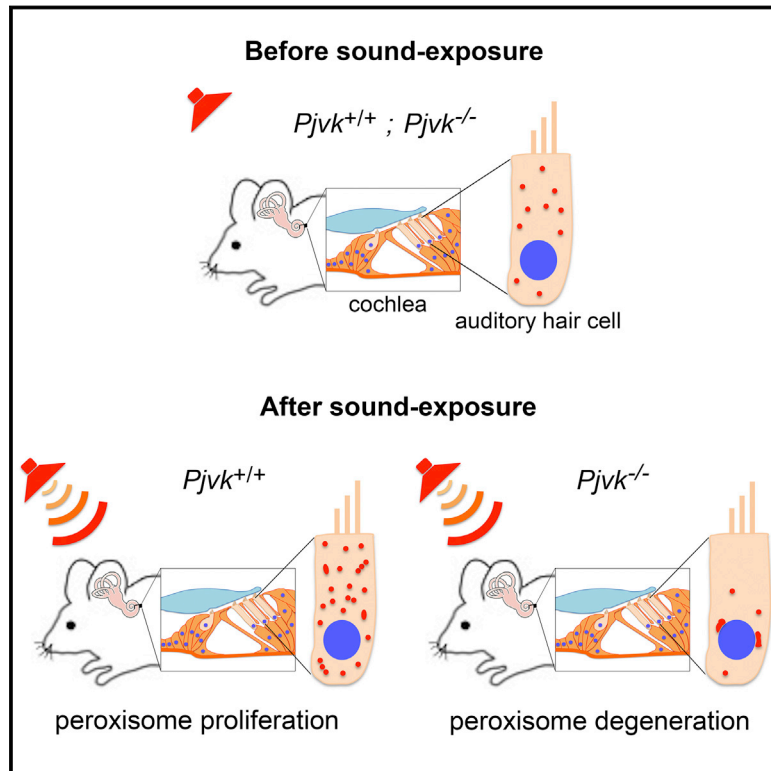
REFERENCES

Agmon, N., Liefshitz, B., Zimmer, C., Fabre, E., and Kupiec, M. (2013). Effect of nuclear architecture on the efficiency of double-strand break repair. Nat. Cell Biol. 15, 694–699.

Aten, J.A., Stap, J., Krawczyk, P.M., van Oven, C.H., Hoebe, R.A., Essers, J., and Kanaar, R. (2004). Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. Science 303, 92–95.

Banerjee, S., Kaye, S.B., and Ashworth, A. (2010). Making the best of PARP inhibitors in ovarian cancer. Nat. Rev. Clin. Oncol. 7, 508–519.

Becker, A., Durante, M., Taucher-Scholz, G., and Jakob, B. (2014). ATM alters the otherwise robust chromatin mobility at sites of DNA double-strand breaks (DSBs) in human cells. PLoS ONE 9, e92640.

Boersma, V., Moatti, N., Segura-Bayona, S., Peuscher, M.H., van der Torre, J., Wevers, B.A., Orthwein, A., Durocher, D., and Jacobs, J.J. (2015). MAD2L2 controls DNA repair at telomeres and DNA breaks by inhibiting 5′ end resection. Nature 521, 537–540.

Bothmer, A., Robbiani, D.F., Di Virgilio, M., Bunting, S.F., Klein, I.A., Feldhahn, N., Barlow, J., Chen, H.T., Bosque, D., Callen, E., et al. (2011). Regulation of DNA end joining, resection, and immunoglobulin class switch recombination by 53BP1. Mol. Cell 42, 319–329.

Bouwman, P., Aly, A., Escandell, J.M., Pieterse, M., Bartkova, J., van der Gulden, H., Hiddingh, S., Thanasoula, M., Kulkarni, A., Yang, Q., et al. (2010). 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. Nat. Struct. Mol. Biol. 17, 688–695.

Bunting, S.F., Callén, E., Wong, N., Chen, H.T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L., et al. (2010). 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. Cell *141*, 243–254.

Callen, E., Di Virgilio, M., Kruhlak, M.J., Nieto-Soler, M., Wong, N., Chen, H.T., Faryabi, R.B., Polato, F., Santos, M., Starnes, L.M., et al. (2013). 53BP1 mediates productive and mutagenic DNA repair through distinct phosphoprotein interactions. Cell *153*, 1266–1280.

Cao, L., Xu, X., Bunting, S.F., Liu, J., Wang, R.H., Cao, L.L., Wu, J.J., Peng, T.N., Chen, J., Nussenzweig, A., et al. (2009). A selective requirement for 53BP1 in the biological response to genomic instability induced by Brca1 deficiency. Mol. Cell *35*, 534–541.

Chang, W., Worman, H.J., and Gundersen, G.G. (2015). Accessorizing and anchoring the LINC complex for multifunctionality. J. Cell Biol. *208*, 11–22.

Chapman, J.R., Barral, P., Vannier, J.B., Borel, V., Steger, M., Tomas-Loba, A., Sartori, A.A., Adams, I.R., Batista, F.D., and Boulton, S.J. (2013). RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. Mol. Cell *49*, 858–871.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491.

Cho, N.W., Dilley, R.L., Lampson, M.A., and Greenberg, R.A. (2014). Interchromosomal homology searches drive directional ALT telomere movement and synapsis. Cell *159*, 108–121.

Crisp, M., Liu, Q., Roux, K., Rattner, J.B., Shanahan, C., Burke, B., Stahl, P.D., and Hodzic, D. (2006). Coupling of the nucleus and cytoplasm: role of the LINC complex. J. Cell Biol. *172*, 41–53.

Di Virgilio, M., Callen, E., Yamane, A., Zhang, W., Jankovic, M., Gitlin, A.D., Feldhahn, N., Resch, W., Oliveira, T.Y., Chait, B.T., et al. (2013). Rif1 prevents resection of DNA breaks and promotes immunoglobulin class switching. Science *339*, 711–715.

Dimitrova, N., Chen, Y.C., Spector, D.L., and de Lange, T. (2008). 53BP1 promotes non-homologous end joining of telomeres by increasing chromatin mobility. Nature *456*, 524–528.

Ding, X., Xu, R., Yu, J., Xu, T., Zhuang, Y., and Han, M. (2007). SUN1 is required for telomere attachment to nuclear envelope and gametogenesis in mice. Dev. Cell *12*, 863–872.

Dion, V., and Gasser, S.M. (2013). Chromatin movement in the maintenance of genome stability. Cell *152*, 1355–1364.

Dion, V., Kalck, V., Horigome, C., Towbin, B.D., and Gasser, S.M. (2012). Increased mobility of double-strand breaks requires Mec1, Rad9 and the homologous recombination machinery. Nat. Cell Biol. *14*, 502–509.

Escribano-Díaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J.T., Tkáč, J., Cook, M.A., Rosebrock, A.P., Munro, M., Canny, M.D., et al. (2013). A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. Mol. Cell *49*, 872–883.

Falk, M., Lukasova, E., Gabrielova, B., Ondrej, V., and Kozubek, S. (2007). Chromatin dynamics during DSB repair. Biochim. Biophys. Acta *1773*, 1534–1545.

Feng, L., Fong, K.W., Wang, J., Wang, W., and Chen, J. (2013). RIF1 counteracts BRCA1-mediated end resection during DNA repair. J. Biol. Chem. *288*, 11135–11143.

Fumagalli, M., Rossiello, F., Clerici, M., Barozzi, S., Cittaro, D., Kaplunov, J.M., Bucci, G., Dobreva, M., Matti, V., Beausejour, C.M., et al. (2012). Telomeric DNA damage is irreparable and causes persistent DNA-damage-response activation. Nat. Cell Biol. *14*, 355–365.

Gandhi, M., Evdokimova, V.N.T, T Cuenco, K., Nikiforova, M.N., Kelly, L.M., Stringer, J.R., Bakkenist, C.J., and Nikiforov, Y.E. (2012). Homologous chromosomes make contact at the sites of double-strand breaks in genes in somatic G0/G1-phase human cells. Proc. Natl. Acad. Sci. USA *109*, 9454–9459.

Jakob, B., Splinter, J., and Taucher-Scholz, G. (2009). Positional stability of damaged chromatin domains along radiation tracks in mammalian cells. Radiat. Res. *171*, 405–418.

Ketema, M., Wilhelmsen, K., Kuikman, I., Janssen, H., Hodzic, D., and Sonnenberg, A. (2007). Requirements for the localization of nesprin-3 at the nuclear envelope and its interaction with plectin. J. Cell Sci. *120*, 3384–3394.

Kruhlak, M.J., Celeste, A., Dellaire, G., Fernandez-Capetillo, O., Müller, W.G., McNally, J.G., Bazett-Jones, D.P., and Nussenzweig, A. (2006). Changes in chromatin structure and mobility in living cells at sites of DNA double-strand breaks. J. Cell Biol. *172*, 823–834.

Lei, K., Zhang, X., Ding, X., Guo, X., Chen, M., Zhu, B., Xu, T., Zhuang, Y., Xu, R., and Han, M. (2009). SUN1 and SUN2 play critical but partially redundant roles in anchoring nuclei in skeletal muscle cells in mice. Proc. Natl. Acad. Sci. USA *106*, 10207–10212.

Lottersberger, F., Bothmer, A., Robbiani, D.F., Nussenzweig, M.C., and de Lange, T. (2013). Role of 53BP1 oligomerization in regulating double-strand break repair. Proc. Natl. Acad. Sci. USA *110*, 2146–2151.

Mahen, R., Hattori, H., Lee, M., Sharma, P., Jeyasekharan, A.D., and Venkitaraman, A.R. (2013). A-type lamins maintain the positional stability of DNA damage repair foci in mammalian nuclei. PLoS ONE *8*, e61893.

Miné-Hattab, J., and Rothstein, R. (2012). Increased chromosome mobility facilitates homology search during recombination. Nat. Cell Biol. *14*, 510–517.

Munoz, I.M., Jowsey, P.A., Toth, R., and Rouse, J. (2007). Phospho-epitope binding by the BRCT domains of hPTIP controls multiple aspects of the cellular response to DNA damage. Nucleic Acids Res. *35*, 5312–5322.

Neumaier, T., Swenson, J., Pham, C., Polyzos, A., Lo, A.T., Yang, P., Dyball, J., Asaithamby, A., Chen, D.J., Bissell, M.J., et al. (2012). Evidence for formation of DNA repair centers and dose-response nonlinearity in human cells. Proc. Natl. Acad. Sci. USA *109*, 443–448.

Padmakumar, V.C., Libotte, T., Lu, W., Zaim, H., Abraham, S., Noegel, A.A., Gotzmann, J., Foisner, R., and Karakesisoglou, I. (2005). The inner nuclear membrane protein Sun1 mediates the anchorage of Nesprin-2 to the nuclear envelope. J. Cell Sci. *118*, 3419–3430.

Palm, W., and de Lange, T. (2008). How shelterin protects mammalian telomeres. Annu. Rev. Genet. *42*, 301–334.

Panier, S., and Boulton, S.J. (2014). Double-strand break repair: 53BP1 comes into focus. Nat. Rev. Mol. Cell Biol. *15*, 7–18.

Peterson, C.L., and Almouzni, G. (2013). Nucleosome dynamics as modular systems that integrate DNA damage and repair. Cold Spring Harb. Perspect. Biol. *5*, a012658.

Rothkamm, K., and Löbrich, M. (2003). Evidence for a lack of DNA double-strand break repair in human cells exposed to very low x-ray doses. Proc. Natl. Acad. Sci. USA *100*, 5057–5062.

Roux, K.J., Crisp, M.L., Liu, Q., Kim, D., Kozlov, S., Stewart, C.L., and Burke, B. (2009). Nesprin 4 is an outer nuclear membrane protein that can induce kinesin-mediated cell polarization. Proc. Natl. Acad. Sci. USA *106*, 2194–2199.

Sato, A., Isaac, B., Phillips, C.M., Rillo, R., Carlton, P.M., Wynne, D.J., Kasad, R.A., and Dernburg, A.F. (2009). Cytoskeletal forces span the nuclear envelope to coordinate meiotic chromosome pairing and synapsis. Cell *139*, 907–919.

Sbalzarini, I.F., and Koumoutsakos, P. (2005). Feature point tracking and trajectory analysis for video imaging in cell biology. J. Struct. Biol. *151*, 182–195.

Seeber, A., Dion, V., and Gasser, S.M. (2013). Checkpoint kinases and the INO80 nucleosome remodeling complex enhance global chromatin mobility in response to DNA damage. Genes Dev. *27*, 1999–2008.

Shibuya, H., and Watanabe, Y. (2014). The meiosis-specific modification of mammalian telomeres. Cell Cycle *13*, 2024–2028.

Soutoglou, E., Dorn, J.F., Sengupta, K., Jasin, M., Nussenzweig, A., Ried, T., Danuser, G., and Misteli, T. (2007). Positional stability of single double-strand breaks in mammalian cells. Nat. Cell Biol. *9*, 675–682.

Starr, D.A., and Fridolfsson, H.N. (2010). Interactions between nuclei and the cytoskeleton are mediated by SUN-KASH nuclear-envelope bridges. Annu. Rev. Cell Dev. Biol. *26*, 421–444.

Swartz, R.K., Rodriguez, E.C., and King, M.C. (2014). A role for nuclear enve-lope-bridging complexes in homology-directed repair. Mol. Biol. Cell 25, 2461–2471.

Thévenaz, P., Ruttimann, U.E., and Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. IEEE Trans. Image Process. 7, 27–41.

Turgay, Y., Champion, L., Balazs, C., Held, M., Toso, A., Gerlich, D.W., Mer-aldi, P., and Kutay, U. (2014). SUN proteins facilitate the removal of mem-branes from chromatin during nuclear envelope breakdown. J. Cell Biol. 204, 1099–1109.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. Science 343, 80–84.

Wilson, K.L., and Foisner, R. (2010). Lamin-binding Proteins. Cold Spring Harb. Perspect. Biol. 2, a000554.

Wu, P., van Overbeek, M., Rooney, S., and de Lange, T. (2010). Apollo contrib-utes to G overhang maintenance and protects leading-end telomeres. Mol. Cell 39, 606–617.

Xu, G., Chapman, J.R., Brandsma, I., Yuan, J., Mistrik, M., Bouwman, P., Bart-kova, J., Gogola, E., Warmerdam, D., Barazas, M., et al. (2015). REV7 counter-acts DNA double-strand break resection and affects PARP inhibition. Nature 521, 541–544.

Zidovska, A., Weitz, D.A., and Mitchison, T.J. (2013). Micron-scale coherence in interphase chromatin dynamics. Proc. Natl. Acad. Sci. USA 110, 15555–15560.

Zimmermann, M., and de Lange, T. (2014). 53BP1: pro choice in DNA repair. Trends Cell Biol. 24, 108–117.

Zimmermann, M., Lottersberger, F., Buonomo, S.B., Sfeir, A., and de Lange, T. (2013). 53BP1 regulates DSB repair using Rif1 to control 5′ end resection. Sci-ence 339, 700–704.

Cell

Hypervulnerability to Sound Exposure through Impaired Adaptive Proliferation of Peroxisomes

Graphical Abstract

Authors

Sedigheh Delmaghani, Jean Defourny, Asadollah Aghaie, ..., Jean-Pierre Hardelin, Paul Avan, Christine Petit

Correspondence

christine.petit@pasteur.fr

In Brief

Hypervulnerability to sound exposure in pejvakin-deficient mice results from impaired adaptive proliferation of peroxisomes in response to the oxidative stress. In wild-type mice, loud sounds elicit adaptive peroxisomal proliferation.

Highlights

- Pejvakin-deficient mice and humans are hypervulnerable to sound exposure

- Oxidative stress induces a pejvakin-dependent proliferation of peroxisomes

- Peroxisome proliferation contributes to the physiological response to sound exposure

- *Pjvk* gene transfer can rescue auditory dysfunction in *Pjvk*$^{-/-}$ mice

Accession Number

GSE72722

CrossMark

CellPress

Hypervulnerability to Sound Exposure through Impaired Adaptive Proliferation of Peroxisomes

Sedigheh Delmaghani,[1,2,3] Jean Defourny,[1,2,3] Asadollah Aghaie,[2,3,4] Maryline Beurg,[5] Didier Dulon,[5] Nicolas Thelen,[6] Isabelle Perfettini,[1,2,3] Tibor Zelles,[7,8] Mate Aller,[7] Anaïs Meyer,[1,2,3] Alice Emptoz,[1,2,3] Fabrice Giraudet,[9,10,11] Michel Leibovici,[1,2,3] Sylvie Dartevelle,[12] Guillaume Soubigou,[13] Marc Thiry,[6] E. Sylvester Vizi,[7] Saaid Safieddine,[1,2,3] Jean-Pierre Hardelin,[1,2,3] Paul Avan,[9,10,11,15] and Christine Petit[1,2,3,4,14,15,*]

[1]Unité de Génétique et Physiologie de l'Audition, Institut Pasteur, 75015 Paris, France
[2]UMRS 1120, Institut National de la Santé et de la Recherche Médicale (INSERM), 75015 Paris, France
[3]Sorbonne Universités, UPMC Université Paris 06, Complexité du Vivant, 75005 Paris, France
[4]Syndrome de Usher et Autres Atteintes Rétino-Cochléaires, Institut de la Vision, 75012 Paris, France
[5]Equipe Neurophysiologie de la Synapse Auditive, Université de Bordeaux, Neurosciences Institute, CHU Pellegrin, 33076 Bordeaux, France
[6]Unit of Cell and Tissue Biology, GIGA-Neurosciences, University of Liege, CHU Sart-Tilman, B36, 4000 Liege, Belgium
[7]Institute of Experimental Medicine, Hungarian Academy of Sciences, 1083 Budapest, Hungary
[8]Department of Pharmacology and Pharmacotherapy, Semmelweis University, 1089 Budapest, Hungary
[9]Laboratoire de Biophysique Sensorielle, Université d'Auvergne, 63000 Clermont-Ferrand, France
[10]UMR 1107, Institut National de la Santé et de la Recherche Médicale (INSERM), 63000 Clermont-Ferrand, France
[11]Centre Jean Perrin, 63000 Clermont-Ferrand, France
[12]Plateforme d'Ingénierie des Anticorps, Institut Pasteur, 75015 Paris, France
[13]Plateforme Transcriptome et Épigénome, Institut Pasteur, 75015 Paris, France
[14]Collège de France, 75005 Paris, France
[15]Co-senior author
*Correspondence: christine.petit@pasteur.fr
http://dx.doi.org/10.1016/j.cell.2015.10.023

SUMMARY

A deficiency in pejvakin, a protein of unknown function, causes a strikingly heterogeneous form of human deafness. Pejvakin-deficient ($Pjvk^{-/-}$) mice also exhibit variable auditory phenotypes. Correlation between their hearing thresholds and the number of pups per cage suggest a possible harmful effect of pup vocalizations. Direct sound or electrical stimulation show that the cochlear sensory hair cells and auditory pathway neurons of $Pjvk^{-/-}$ mice and patients are exceptionally vulnerable to sound. Subcellular analysis revealed that pejvakin is associated with peroxisomes and required for their oxidative-stress-induced proliferation. $Pjvk^{-/-}$ cochleas display features of marked oxidative stress and impaired antioxidant defenses, and peroxisomes in $Pjvk^{-/-}$ hair cells show structural abnormalities after the onset of hearing. Noise exposure rapidly upregulates Pjvk cochlear transcription in wild-type mice and triggers peroxisome proliferation in hair cells and primary auditory neurons. Our results reveal that the antioxidant activity of peroxisomes protects the auditory system against noise-induced damage.

INTRODUCTION

Mutations of *PJVK*, which encodes pejvakin, a protein of unknown function present only in vertebrates, cause the DFNB59-recessive form of sensorineural hearing impairment. In the first patients described (Delmaghani et al., 2006), the impairment was restricted to neurons of the auditory pathway, with auditory brainstem responses (ABRs) displaying abnormally decreased wave amplitudes and increased inter-wave latencies (Starr and Rance, 2015). ABRs monitor the electrical response of auditory pathways to brief sound stimuli, from the primary auditory neurons synapsing with the sensory cells of the cochlea, the inner hair cells (IHCs), to the colliculus in the midbrain (Møller and Jannetta, 1983). However, some DFNB59 patients were found to have a cochlear dysfunction, as shown by an absence of the otoacoustic emissions (OAEs) that are produced by the outer hair cells (OHCs), frequency-tuned cells endowed with electromotility that mechanically amplify the sound stimulation of neighboring IHCs (Ashmore, 2008). These patients had truncating mutations of *PJVK*, whereas the previously identified patients, with extant OAEs, had missense mutations (p.T54I or p.R183W) (Ebermann et al., 2007; Schwander et al., 2007; Borck et al., 2012). However, the identification of patients also carrying the p.R183W missense mutation but lacking OAEs (Collin et al., 2007) refuted any straightforward connection between the nature of the *PJVK* mutation and the hearing phenotype. The severity of deafness in DFNB59 patients varies from moderate to profound
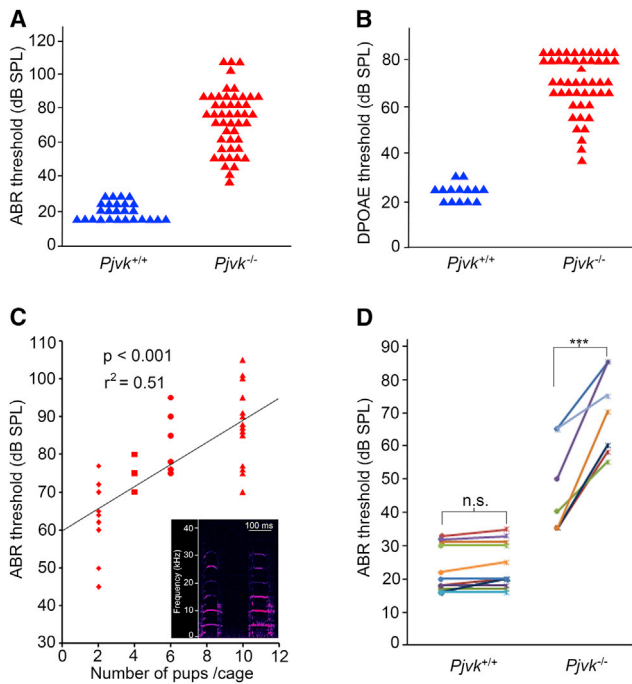
Figure 1. Hearing Loss Variability and Greater Sensitivity to Controlled Sound Exposure in Pjvk$^{-/-}$ Mice

(A) ABR thresholds at 10 kHz in P30 Pjvk$^{+/+}$ (n = 26 mice) and Pjvk$^{-/-}$ (n = 48 mice) littermates.

(B) DPOAE thresholds at 10 kHz in P30 Pjvk$^{+/+}$ (n = 14 mice) and Pjvk$^{-/-}$ (n = 48 mice) littermates. In ears with no DPOAE, even at 75 dB SPL (the highest sound intensity tested), DPOAE thresholds were arbitrarily set at 80 dB SPL.

(C) Relationship between the number of pups raised together (determining sound levels in the immediate environment) and ABR thresholds at 10 kHz in P21 Pjvk$^{-/-}$ pups. Inset: a time-frequency analysis of a mouse pup's vocalization. Pup calls from P0 to P21 form harmonic series of about 5 kHz, with the most energetic harmonic at about 10 kHz. In a 12-pup litter, call levels reach 105 ± 5 dB SPL at the entrance to the ear canals of the pups.

(D) ABR thresholds at 10 kHz in P30 Pjvk$^{+/+}$ and Pjvk$^{-/-}$ mice before (dots) and after (crosses) controlled sound exposure. ns, not significant; ***p < 0.001. See also Figure S1.

and may even be progressive in some patients, suggesting that extrinsic factors may influence the hearing phenotype.

We investigated the role of pejvakin, with the aim of determining the origin of the phenotypic variability of the DFNB59 form of deafness. Our study of Pjvk knockout mouse models and of patients revealed an unprecedented hypervulnerability of auditory hair cells and neurons to sound exposure, accounting for phenotypic variability. We found that pejvakin is a peroxisome-associated protein involved in the oxidative-stress-induced proliferation of this organelle. Pejvakin-deficient mice revealed the key role of peroxisomes in the redox homeostasis of the auditory system and in the protection against noise-induced hearing loss.

RESULTS

Heterogeneity in the Hearing Sensitivity of Pjvk$^{-/-}$ Mice

We generated pejvakin-null (Pjvk$^{-/-}$) mice carrying a deletion of Pjvk exon 2, resulting in a frameshift at codon position 71

(p.Gly71fs*9) (Figure S1; see the Supplemental Experimental Procedures). ABR thresholds recorded on postnatal day 30 (P30) Pjvk$^{-/-}$ mice (n = 48) ranged from 35 to 110 dB SPL (sound pressure level) at 10 kHz but never exceeded 30 dB SPL in their Pjvk$^{+/+}$ littermates (n = 26) (Figure 1A). This broad range of hearing sensitivity in Pjvk$^{-/-}$ mice, from near-normal hearing to almost complete deafness, extended across the whole frequency spectrum. The thresholds of distortion-product OAEs (DPOAEs) at 10 kHz (i.e., the minimum stimulus required for DPOAEs production by OHCs) also fell within an abnormally large range of values, from 30 to 75 dB SPL, in 28 Pjvk$^{-/-}$ mice, indicating an OHC dysfunction, and DPOAEs were undetectable in another 20 Pjvk$^{-/-}$ mice, suggesting a complete OHC defect (Figure 1B). The absence of pejvakin in mice thus results in a puzzlingly large degree of hearing phenotype variability.

Hypervulnerability to the Natural Acoustic Environment in Pjvk$^{-/-}$ Mice

We investigated the variability of Pjvk$^{-/-}$ auditory phenotypes, by first determining the ABR thresholds of Pjvk$^{-/-}$ littermates from different crosses. Large differences were observed between crosses, with much fewer differences between the Pjvk$^{-/-}$ littermates of individual crosses. Litters with larger numbers of pups (6 to 12) had higher ABR thresholds, suggesting that the natural acoustic environment, with the calls of larger numbers of pups, might be deleterious in Pjvk$^{-/-}$ mice. Pups are vocally very active from birth to about P20. We manipulated the level of exposure to pup calls by randomly splitting large litters of Pjvk$^{-/-}$ pups into groups of 2, 4, 6 and 10 pups per cage, with foster mothers, before P10, i.e., several days before hearing onset. The ABR thresholds at P21 were significantly correlated with the number of pups raised together (p < 0.001, r^2 = 0.51) (Figure 1C).

We then evaluated the effect of a controlled sound stimulation on hearing, by presenting 1,000 tone bursts at 10 kHz, 105 dB SPL (2-ms plateau stimulations separated by 60-ms intervals of silence), energetically equivalent to a 3-min stay in the natural environment of a 12-pup litter, while monitoring the ABRs during sound exposure. These conditions are referred to hereafter as "controlled sound exposure." We probed the effect of sound exposure by ABR tests, which, limited to 50 repetitions of tone bursts, did not influence the hearing thresholds of Pjvk$^{-/-}$ mice. In a sample of P30 Pjvk$^{-/-}$ mice with initial ABR threshold elevation (below 35 dB SPL), controlled sound exposure affected ABR thresholds in the 12–20 kHz frequency interval (corresponding to the cochlear zones in which hair-cell stimulation was strongest), with an immediate increase of 21.7 ± 10.3 dB (n = 8; p < 0.001), not observed in Pjvk$^{+/+}$ mice (2.2 ± 2.4 dB, n = 12; p = 0.3) (Figure 1D). Pjvk$^{-/-}$ mice transferred to a silent environment after exposure displayed a further increase of 33.7 ± 16.0 dB (n = 8) 2 days after exposure. The threshold shift decreased to 23.7 ± 18.0 dB at 7 days, and disappeared entirely by 14 days. When exposed mice were returned to the box with their littermates, their ABR continued to increase, at a rate of 15 dB per week. Pejvakin deficiency thus results in particularly high levels of vulnerability to low levels of acoustic energy, and the increase in ABR thresholds is reversible but only slowly and in a quiet environment.

Hair Cells and Auditory Pathway Neurons Are Affected by Pejvakin Deficiency

To identify the cellular targets of the pejvakin deficiency, we specifically probed the function of auditory hair cells and neurons in $Pjvk^{-/-}$, hair cell-conditional $Pjvk$ knockout ($Pjvk^{fl/fl}$Myo15-$cre^{+/-}$), and $Pjvk^{+/+}$ mice, at the age of 3 weeks, before and after controlled sound exposure or controlled electrical stimulation. The responses of the IHCs to sound-induced vibrations amplified by OHCs trigger action potentials in the distal part of primary auditory neurons, at the origin of ABR wave I. In $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice, which lack pejvakin only in the hair cells, ABR wave I amplitude and latency at 105 dB SPL specifically probed IHC function, because IHC responses to such loud sounds are independent of OHC activity (Robles and Ruggero, 2001). The larger wave I latency (1.58 ms in $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice [n = 20] versus 1.32 ms in $Pjvk^{+/+}$ littermates [n = 30]; p < 0.001) and lower wave I amplitude (37% of the amplitude in $Pjvk^{+/+}$ littermates; p < 0.001) suggested a dysfunction of the IHCs. Controlled sound exposure induced further decreases in ABR wave I amplitude in $Pjvk^{-/-}$ and $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice (48% and 55% of pre-exposure amplitude, respectively) with respect to $Pjvk^{+/+}$ mice (108%; p < 0.001 for both comparisons) (Figure 2A), demonstrating that $Pjvk^{-/-}$ IHCs are hypervulnerable to sound. As shown above, OHCs are also affected by the pejvakin deficiency. Controlled sound exposure triggered a mean decrease in the DPOAE amplitude of 16.9 ± 7.2 dB in the 12 to 20 kHz frequency interval in $Pjvk^{-/-}$ mice with persistent DPOAEs (n = 8; p < 0.0001), and an increase in DPOAE threshold, but it had no effect on the DPOAEs of $Pjvk^{+/+}$ mice (n = 9; p = 0.51) (Figure 2B). OHCs lacking pejvakin are thus also hypervulnerable to sound.

We investigated the effect of the absence of pejvakin on the auditory pathway by comparing electrically evoked brainstem responses (EEBR) in $Pjvk^{-/-}$ and $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice (see the Supplemental Experimental Procedures). The amplitudes of the most distinctive EEBR waves, E II and E IV, did not differ between the two types of mice (for wave E IV: 2.6 ± 1.8 μV in $Pjvk^{-/-}$ mice [n = 18] and 2.2 ± 1.2 μV in $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice [n = 11]; t test, p = 0.13). However, following controlled electrical exposure at 200 impulses/s for 1 min, as opposed to electric-impulse stimulation with 16 impulses/s for 10 s for pre- and post-exposure EEBR tests, E II and E IV EEBR wave amplitudes got 41% and 47% smaller, respectively, for at least 3 min, in $Pjvk^{-/-}$ mice (n = 5; paired t test, p = 0.02 and p = 0.01, respectively), but were unaffected in $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice (n = 10; p = 0.83) (Figures 2D and 2G–2I). The E II-E IV interwave interval was 0.41 ms longer in $Pjvk^{-/-}$ mice (n = 5) than in $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice (n = 10; p = 0.003), and controlled electrical exposure extended this interval by a further 0.15 ms in $Pjvk^{-/-}$ mice only (paired t test, p = 0.001) (Figures 2H and 2I). Likewise, the latency interval between ABR wave I and wave IV (the counterpart of wave E IV), abnormal in one-third of the $Pjvk^{-/-}$ mice tested (with an ABR threshold < 95 dB SPL, n = 12) (Figures 2C and 2E), got abnormal in all of them after controlled sound exposure (0.16 ms further increase; paired t test, p < 0.001). By contrast, it remained normal in $Pjvk^{fl/fl}$Myo15-$cre^{+/-}$ mice (n = 10 ears; p = 0.73) (Figures 2C and 2F). Thus, the absence of pejvakin affects the propagation

of action potentials in the auditory pathway after both controlled electrical and sound exposure in the $Pjvk^{-/-}$ mice.

To clarify whether these abnormalities were of neuronal or glial origin, we performed a rescue experiment in $Pjvk^{-/-}$ mice, using adeno-associated virus 8 (AAV8) vector-mediated transfer of the murine pejvakin cDNA (AAV8-Pjvk). AAV8 injected into the cochlea transduces the primary auditory neurons (cochlear ganglion neurons) and neurons of the cochlear nucleus (Figure S2A), but not the hair cells. All $Pjvk^{-/-}$ mice (n = 7) injected on P3 and tested on P21 had normal ABR interwave I-IV latencies (Figure 2J), and their EEBR wave-E IV amplitude was insensitive to controlled electrical stimulation (1.91 ± 0.97 μV before and 1.87 μV ± 1.07 after stimulation; paired t test, p = 0.59) (Figures 2K and 2L). The absence of pejvakin thus renders auditory pathway neurons hypervulnerable to exposure to mild, short sound stimuli.

Hypervulnerability to Sound in DFNB59 Patients

We then investigated whether the hearing of DFNB59 patients was also hypervulnerable to sound exposure. We tested five patients carrying the p.T54I mutation (Delmaghani et al., 2006). Transient-evoked OAEs (TEOAEs) assessing OHC function over a broad range of frequencies were detected for all ears, despite the severe hearing impairment (hearing threshold increasing from 66 dB HL at 250 Hz to 84 dB at 8 kHz). Following minimal exposure to impulse stimuli (clicks at 99 dB nHL), ABR waves were clearly identified in response to 250 clicks. When exposure was prolonged to 1,000 clicks (the standard procedure), wave V, the equivalent of mouse ABR-wave IV, which was initially conspicuous, displayed a decrease in amplitude (to 39% ± 30% of its initial amplitude) and an increase in latency (of 0.30 ± 0.15 ms) (Figures 3A, 3C, and 3D). In parallel, the I-V interwave interval increased by 0.30 ± 0.15 ms. Wave-V amplitude and latency recovered fully after 10 min of silence (Figure 3B). In control patients with sensorineural hearing impairment of cochlear origin matched for ABR thresholds, similar sound stimulation did not affect ABR wave-V amplitude (105% ± 14% of the initial amplitude after exposure; n = 13 patients) or latency (−0.02 ± 0.07 ms change after exposure) (Figures 3C and 3D). Exposure of the DFNB59 patients to 1,000 clicks also affected TEOAEs (6.1 ± 5.2 dB nHL decrease in amplitude; paired t test, p = 0.02). Therefore, as in pejvakin-deficient mice, the cochlear and neuronal responses of DFNB59 patients were affected by exposure to low-energy sound.

Redox Status Abnormalities and ROS-Induced Cell Damage in the Cochlea of $Pjvk^{-/-}$ Mice

We studied the $Pjvk^{-/-}$ cochlea by light microscopy on semithin sections and electron microscopy. On P15 and P21, both OHCs and IHCs were normal in number and shape. Their hair bundles (the mechanoreceptive structures responding to sound), the ribbon synapses of the IHCs, and their primary auditory neurons were unmodified (data not shown). On P30, we observed the loss of a few OHCs (16% ± 11%, n = 5 mice), restricted to the basal region of the cochlea (tuned to high-frequency sounds). From P30 onward, OHCs, cochlear ganglion neurons, and then IHCs disappeared, and the sensory epithelium (organ of Corti) progressively degenerated (Figure S3A).
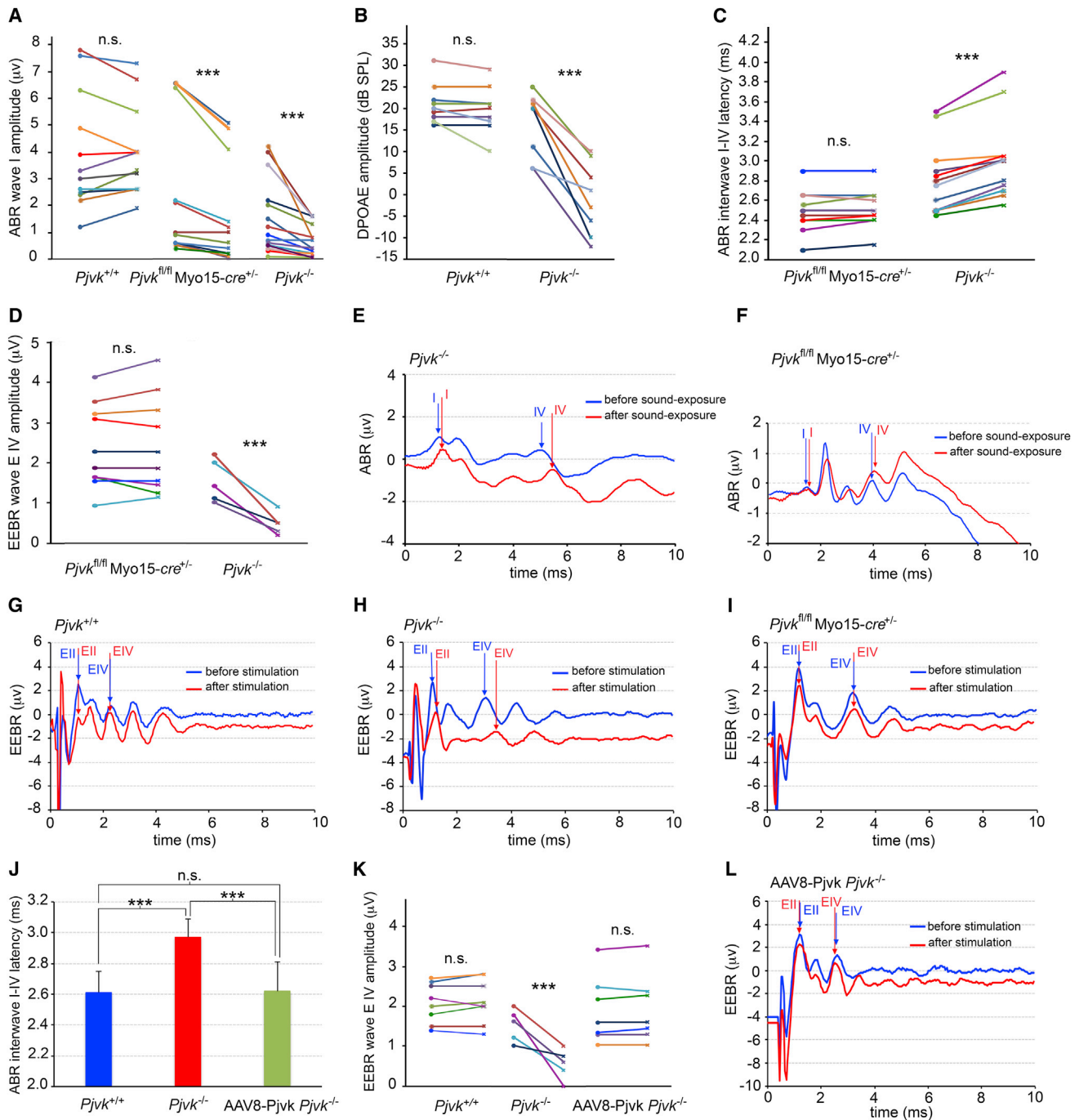
Figure 2. Effects on Auditory Function of Brief Exposure to Moderately Intense Stimuli in *Pjvk*[+/+], *Pjvk*[−/−], and *Pjvk*[fl/fl]Myo15-*cre*[+/−] Mice

(A–C) ABR wave I amplitude (A), DPOAE amplitude (B), and ABR interwave I-IV latency (C) in *Pjvk*[+/+], *Pjvk*[−/−], and *Pjvk*[fl/fl]Myo15-*cre*[+/−] mice, before (dots) and after (crosses) controlled sound exposure, revealing the hypervulnerability to sound of both types of cochlear hair cells (IHCs and OHCs) and of the neural pathway.

(D) EEBR wave E IV amplitude before and after controlled electrical exposure in *Pjvk*[−/−] and *Pjvk*[fl/fl]Myo15-*cre*[+/−] mice was abnormal and hypervulnerable only when pejvakin is absent from auditory neurons (*Pjvk*[−/−] mice).

(E and F) Examples of ABRs in *Pjvk*[−/−] and *Pjvk*[fl/fl]Myo15-*cre*[+/−] mice: the latency of wave I is affected by controlled sound exposure in both mutant mice, and wave IV displays an additional increase in latency only in *Pjvk*[−/−] mice.

(G–I) Examples of EEBRs in *Pjvk*[+/+] (G), *Pjvk*[−/−] (H), and *Pjvk*[fl/fl]Myo15-*cre*[+/−] (I) mice; EEBRs are affected by controlled electrical exposure only in *Pjvk*[−/−] mice.

(legend continued on next page)

We investigated possible changes in gene expression in the organ of Corti of P15 $Pjvk^{-/-}$ mice, by microarrays (see the Supplemental Experimental Procedures). Eighteen genes had expression levels at least 1.5-fold higher or lower in $Pjvk^{-/-}$ mice than in $Pjvk^{+/+}$ mice. Marked differences were observed for four genes involved in the redox balance—*CypA*, *Gpx2*, *c-Dct*, and *Mpv17*—encoding cyclophilin A, glutathione peroxidase 2, c-dopachrome tautomerase, and Mpv17, respectively (Table S1). All of these genes were downregulated in $Pjvk^{-/-}$ mice, a result confirmed by qRT-PCR (Figure S4A), and all encode antioxidant proteins, suggesting that $Pjvk^{-/-}$ mice have impaired antioxidant defenses (Table S1).

We thus assessed the level of oxidative stress in the cochlea of P21 $Pjvk^{-/-}$ mice, by determining the ratio of reduced to oxidized glutathione (GSH:GSSG). The GSSG content was about three times larger than in $Pjvk^{+/+}$ mice, whereas the GSH content was 23% smaller, resulting in a GSH:GSSG ratio in $Pjvk^{-/-}$ cochleas reduced by a factor of 3.4 (Figure 4A). Pejvakin deficiency thus results in cochlear oxidative stress.

We assessed lipid peroxidation by reactive oxygen species (ROS) in $Pjvk^{-/-}$ mice, by immunofluorescence-based detection of the by-product 4-hydroxy-2-nonenal (4-HNE). Strong immunoreactivity was observed in P60 $Pjvk^{-/-}$ hair cells and cochlear ganglion neurons (Figure S3B). Quantification of lipid peroxidation in microdissected organs of Corti from P30 $Pjvk^{-/-}$ and $Pjvk^{+/+}$ mice, showed a moderate, but statistically significant, increase of the malondialdehyde content in the absence of pejvakin (2.15 ± 0.14 μM in $Pjvk^{-/-}$ versus 1.84 ± 0.11 μM in $Pjvk^{+/+}$ mice; p = 0.04). Thus, pejvakin deficiency led to impaired antioxidant defenses in the cochlea, resulting in ROS-induced cell damage.

We then studied electrophysiological features of IHCs and OHCs in the mature cochlea of P19–P21 $Pjvk^{-/-}$ mice. In IHCs, the number of synaptic ribbons, Ca^{2+} currents, and synaptic exocytosis were unaffected (Figure S5A). We investigated whether $Pjvk^{-/-}$ mice display the main K^+ currents found in mature IHCs, specifically $I_{K,f}$, which plays a major part in IHC repolarization and is involved in the high temporal precision of action potentials in postsynaptic nerve fibers, $I_{K,s}$, and $I_{K,n}$ (Oliver et al., 2006). The $I_{K,f}$ current that flows through the large conductance voltage- and Ca^{2+}-activated potassium (BK) channels, a well-known target of ROS (Tang et al., 2004), was detected in only 4 out of 11 $Pjvk^{-/-}$ IHCs, and the mean number of spots immunolabeled for the BK α-subunit per IHC was much lower in $Pjvk^{-/-}$ mice (5.0 ± 1.4, n = 283 IHCs from seven mice) than in $Pjvk^{+/+}$ mice (13.9 ± 2.6, n = 204 IHCs from nine mice; t test, p < 0.001). By contrast, the $I_{K,s}$ and $I_{K,n}$ currents were not affected (Figures 4B and S5B). The electromotility of OHCs was moderately impaired in $Pjvk^{-/-}$ mice (Figure S5C). This contrasted with the total loss of DPOAE in a large majority of Pjvk$^{-/-}$ mice from P15 on, even at the highest possible stimulus level of 75 dB SPL. It thus pinpointed the existence of an additional defect, likely a mechanoelectrical transduction defect, the main determinant of DPOAEs at high stimulus levels (Avan et al., 2013). The decrease of the cochlear microphonic potential that reflects mechanoelectrical transduction currents through OHCs of the basal-most cochlear region, indeed corroborated the DPOAE measurements: this potential, recorded for a 5-kHz sound stimulus at 95 dB SPL, was always larger than 10 μV in $Pjvk^{+/+}$ mice (n = 8), but fell between 5 and 3 μV in the P21 $Pjvk^{-/-}$ mice with residual DPOAEs (n = 2), and below 1 μV, in the $Pjvk^{-/-}$ mice without persisting DPOAEs (n = 6). Taken together, oxidative stress in the $Pjvk^{-/-}$ cochlea impacts various electrophysiological properties of the hair cells, particularly mechanoelectrical transduction and K^+ current through BK channels.

Mitochondrial defects are a common cause of ROS overproduction. However, we did not find evidence that mitochondria were damaged, as vulnerability of the mitochondrial membrane potential, $\Delta\psi_m$, to the uncoupler carbonyl cyanide 4-(trifluoromethoxy)phenylhydrazone (FCCP) in the organ of Corti and cochlear ganglion was similar in P17-P30 $Pjvk^{-/-}$ and $Pjvk^{+/+}$ mice, and analysis of $Pjvk^{-/-}$ hair cells by transmission electron microscopy (TEM) revealed no mitochondrial abnormalities (Figure S5D; data not shown).

Pejvakin Is a Peroxisome-Associated Protein

By using $Pjvk^{-/-}$ cochlea as control, we found that neither the commercially available antibodies nor our initial polyclonal antibody (Delmaghani et al., 2006) specifically recognized pejvakin (data not shown). Given the limited divergence of the pejvakin amino-acid sequence among vertebrates, we tried to elicit an immune response in $Pjvk^{-/-}$ mice (see the Experimental Procedures). The monoclonal antibody obtained, Pjvk-G21, labeled peroxisomes stained by peroxisome membrane protein 70 (PMP70) antibodies in transfected HeLa cells expressing pejvakin (Figure S6A) and in the human HepG2 hepatoblastoma cell line, which is particularly rich in this organelle (Figure 5A). The specificity of the Pjvk-G21 antibody was demonstrated by the immunolabeling of peroxisomes in the hair cells of $Pjvk^{+/+}$, but not of $Pjvk^{-/-}$ and $Pjvk^{fl/fl}$Myo15-*cre*$^{+/-}$ mice (Figures 5B and S6B).

Prediction programs failed to detect the PTS1 or PTS2 motifs in the pejvakin sequence (Mizuno et al., 2008), the targeting signals for the importation of peroxisomal matrix proteins into the organelle (Smith and Aitchison, 2013), suggesting that pejvakin is a peroxisomal membrane or membrane-associated protein.

Structural Abnormalities of Peroxisomes in the Hair Cells of $Pjvk^{-/-}$ Mice

We investigated the distribution and morphology of peroxisomes by TEM. Peroxisomes were identified on the basis of catalase activity detection using 3,3′-diaminobenzidine as substrate. We focused on OHCs, the first to display a dysfunction in $Pjvk^{-/-}$ mice. On P30, but not on P15, both the distribution and shape of peroxisomes differed between $Pjvk^{-/-}$ and $Pjvk^{+/+}$ mice (Figure 5E). In $Pjvk^{+/+}$ OHCs, the peroxisomes were restricted to an area immediately below the cuticular plate. In $Pjvk^{-/-}$ mice,

---

(J–L) Neuronal function rescue in $Pjvk^{-/-}$ mice by transduction with AAV8-Pjvk: effects on ABR interwave I-IV latency (J), on EEBR wave E IV amplitude and its hypervulnerability to electrical stimulation (K), and on EEBR interwave E II-E IV latency (one example is shown in L, to be compared with H). Vertical arrows indicate the positions of waves I and IV on ABR traces and of waves E II and E IV on EEBR traces. ns, not significant; ***p < 0.001. Error bars represent the SD. See also Figures S1 and S2A.

A
First sound exposure

B
Second sound exposure after 10 min in silence
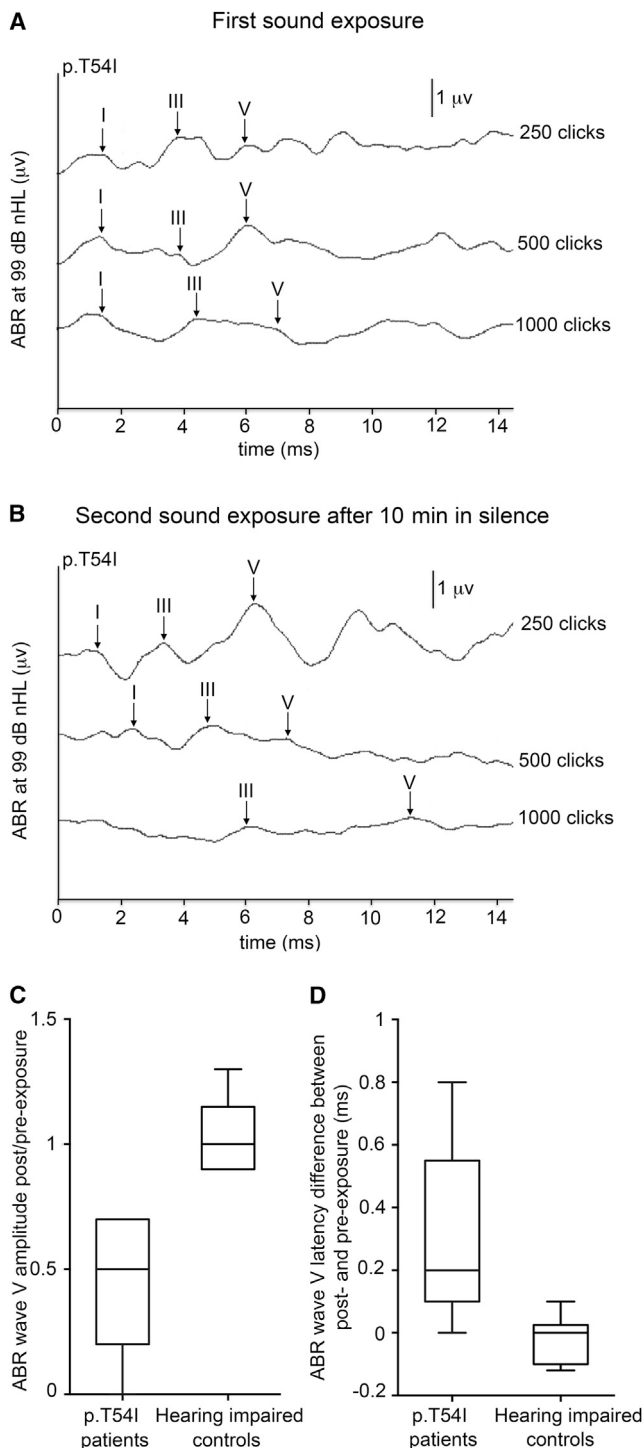

C

D

Figure 3. Hypervulnerability to Sound in DFNB59 Patients

(A) ABR waves I, III, and V (vertical arrows) in one ear of a patient carrying the *PJVK* p.T54I mutation, in response to 250, 500, and 1000 impulse stimuli (clicks) at 99 dB nHL.

(B) Repeated ABRs after 10 min of silence, with an even larger vulnerability of waves I, III, and V.

(C and D) Distributions of the amplitude (C) and latency (D) of ABR wave V in the tested sample of p.T54I patients (n = 8 ears), and in a control group of patients

the peroxisomes located just below the cuticular plate were slightly larger than those in *Pjvk*$^{+/+}$ mice. Strikingly, irregular catalase-containing structures, some of which were juxtaposed, were present in the perinuclear region, in the immediate vicinity of the nuclear membrane of all *Pjvk*$^{-/-}$ OHCs, but not of *Pjvk*$^{+/+}$ OHCs (Figure 5E). The lack of pejvakin thus results in peroxisome abnormalities in OHCs after the onset of hearing.

Pejvakin Is Involved in Oxidative Stress-Induced Peroxisome Proliferation

In HepG2 cells, protrusions emerging from some peroxisomes, the first step of peroxisome biogenesis from pre-existing peroxisomes, were immunoreactive for pejvakin. String-of-beads structures corresponding to elongated and constricted peroxisomes, preceding final fission (Smith and Aitchison, 2013), were also pejvakin-immunoreactive, suggesting a role for this protein in peroxisome proliferation (Figure S6C). Peroxisomes actively contribute to cellular redox balance, by producing and scavenging/degrading H_2O_2 through a broad spectrum of oxidases and peroxidases (especially catalase), respectively (Schrader and Fahimi, 2006). Because *Pjvk*$^{-/-}$ mice displayed features of marked oxidative stress in the cochlea, we investigated the possible role of pejvakin in peroxisome proliferation in response to oxidative stress induced by H_2O_2 (Lopez-Huertas et al., 2000). Embryonic fibroblasts derived from *Pjvk*$^{+/+}$ and *Pjvk*$^{-/-}$ mice were exposed to H_2O_2 (see the Supplemental Experimental Procedures). In unexposed cells, the number of peroxisomes was similar between the two genotypes (t test, p = 0.82). After H_2O_2 treatment, it increased by 46% in *Pjvk*$^{+/+}$ fibroblasts (p = 0.004), but remained unchanged in *Pjvk*$^{-/-}$ fibroblasts (p = 0.83), resulting in a statistically significant difference between the two genotypes (p < 0.001) (Figures 5C and S7A).

We then asked whether mutations reported in DFNB59 patients also affect peroxisome proliferation. We assessed the number of peroxisomes in transfected HeLa cells producing EGFP alone, EGFP and murine pejvakin, or EGFP and one of the mutated forms of murine pejvakin carrying the mutations responsible for DFNB59 (p.T54I, p.R183W, p.C343S, or p.V330Lfs*7). Cells producing the non-mutated pejvakin had larger numbers of peroxisomes than cells producing EGFP alone, whereas cells producing any of the mutated forms of pejvakin (mutPjvk-IRES-EGFP) had smaller peroxisome numbers. In addition, many of these cells contained enlarged peroxisomes, a feature typical of peroxisome proliferation disorders (Ebberink et al., 2012) (Figure 5D and S7B). Together, these results strongly suggest that pejvakin is directly involved in the production of new peroxisomes from pre-existing peroxisomes.

Upregulation of *Pjvk* Cochlear Transcription and Peroxisome Proliferation in Response to Sound

We then asked whether pejvakin is involved in the physiological response to sound. We first assessed the transcription of *Pjvk*

(n = 13) with cochlear hearing impairment and matched ABR thresholds, before and after exposure to clicks #250 to #1000. Boxes extend from the 25th to the 75th percentile. Horizontal bars and vertical bars indicate median values and extremes, respectively. Unlike the unaffected controls, all p.T54I patients displayed markedly decreased amplitudes and increased latencies.
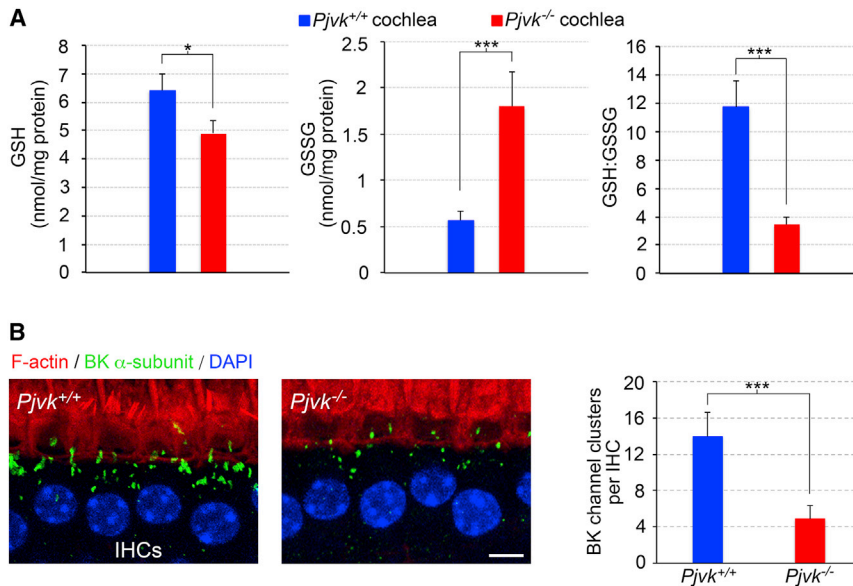
Figure 4. Increased Oxidative Stress and ROS-Induced Cell Damage in the *Pjvk*$^{-/-}$ Cochlea

(A) Reduced glutathione (GSH) (left bar chart), oxidized-glutathione (GSSG) (middle bar chart) contents, and GSH:GSSG ratio (right bar chart) in P21 *Pjvk*$^{-/-}$ versus *Pjvk*$^{+/+}$ cochlea. Error bars represent the SEM of three independent experiments. See also Figure S3.

(B) Marked decrease in the BK α-subunit immunolabeling in *Pjvk*$^{-/-}$ IHCs. Left: P20 *Pjvk*$^{+/+}$ and *Pjvk*$^{-/-}$ IHCs. Scale bar is 5 μm. Right: quantitative analysis of BK channel clusters. Error bars represent the SD. See also Figure S5B. *p < 0.05, ***p < 0.001.

See also Figures S3 and S5.

and of *CypA*, *Gpx2*, *c-Dct*, and *Mpv17*, which were downregulated in *Pjvk*$^{-/-}$ mice, in microdissected organs of Corti from P21 wild-type mice, with or without prior sound stimulation (5–20 kHz, 105 dB SPL for 1 hr; see the Supplemental Experimental Procedures). Transcript levels were analyzed by qRT-PCR at various times (1, 3, 6, and 18 hr) after sound exposure (Figure 6A). *Pjvk* transcript levels had increased by factors of 1.9 ± 0.1 and 3.5 ± 0.7, mean ± SEM, after 1 and 6 hr, respectively. *CypA*, *c-Dct*, and *Mpv17* were also upregulated after 6 hr (by factors of 6.6 ± 1.2, 4.3 ± 0.6, and 1.5 ± 0.1, respectively), as were *c-Fos* and *Hsp70*, used as a positive control, but not *Gpx2*. Thus, noise exposure leads to an upregulation of the transcription of *Pjvk* and of genes downregulated in *Pjvk*$^{-/-}$ mice, and this effect is dependent on acoustic energy level of the stimulation (Figure S4B).

This result predicted that sound exposure would lead to peroxisome proliferation in the auditory system of wild-type mice. 6 hr after exposure (5–20 kHz, 105 dB SPL for 1 hr), the numbers of peroxisomes were unchanged (34.5 ± 0.8 and 35.9 ± 1.0, mean ± SEM, per IHC from unexposed and sound-exposed mice, respectively, n = 75 cells from six mice; t test, p = 0.25). However, at 48 hr, they had markedly increased, by a factor of 2.3, in both IHCs and OHCs (84.7 ± 5.0 per IHC and 16.5 ± 1.0 per OHC, n = 90 cells and n = 150 cells from six mice, respectively) compared to unexposed mice (36.8 ± 3.0 per IHC and 7.3 ± 0.4 per OHC, n = 90 cells and n = 150 cells from six mice, respectively; t test, p < 0.0001 for both comparisons). The number of peroxisomes had also increased, by 35%, in the dendrites of primary auditory neurons (1.7 ± 0.1 and 2.3 ± 0.2 peroxisomes per micrometer of neurite length, n = 40 neurites from five unexposed and five sound-exposed *Pjvk*$^{+/+}$ mice, respectively; t test, p = 0.003) (Figure 6B).

Therapeutic Approaches in *Pjvk*$^{-/-}$ Mice

Based on these results, we tested whether the classical antioxidant drug N-acetyl cysteine (NAC) (either alone or associated with α-lipoic acid and α-tocopherol; see the Supplemental Experimental Procedures) administered to *Pjvk*$^{-/-}$ pups could improve their hearing. The ABR thresholds of P21 NAC-treated *Pjvk*$^{-/-}$ pups (n = 21) were about 10 dB lower than those of untreated *Pjvk*$^{-/-}$ pups (n = 24) for all frequencies tested (t test, p < 0.001 for all comparisons) (Figure 7A). The amplitude of the ABR wave I elicited at 105 dB SPL (4.35 ± 1.16 μV, n = 21) was the same as that of *Pjvk*$^{+/+}$ mice (4.36 ± 1.15 μV, n = 18; t test, p = 0.97) and greater than that of untreated *Pjvk*$^{-/-}$ mice (1.88 ± 1.07 μV, n = 24; t test, p < 0.001) (Figure 7B). EEBRs were more resistant to the high-rate electrical stimulation in treated than in untreated mutant mice (Figure 7C). Conversely, NAC had no beneficial effect on OHCs (data not shown). The association of NAC with α-lipoic acid and α-tocopherol did not perform any better (data not shown).

Full recovery of the neuronal phenotype was achieved by the intracochlear injection of AAV8-Pjvk (see above). As hair cells are not transduced by AAV8, we investigated whether AAV2/8, which transduces hair cells only (Figure S2B), could rescue the *Pjvk*$^{-/-}$ hair-cell phenotype. The auditory function of *Pjvk*$^{-/-}$ mice (n = 7, four pups per cage in every experiment) receiving intracochlear injections of AAV2/8-Pjvk-IRES-EGFP on P3 was assessed on P21, and the percentage of transduced IHCs and OHCs was evaluated in each injected and contralateral (not injected) cochlea, on the basis of EGFP fluorescence. Improvements in ABR thresholds of 20 to 30 dB SPL with respect to untreated mice were observed for frequencies between 10 and 20 kHz (t test, p < 0.001 for all comparisons; Figure 7D). Upon injection of AAV2/8-EGFP, DPOAEs, ABR thresholds, and ABR wave I amplitude and latency were similar to those of untreated *Pjvk*$^{-/-}$ mice (data not shown). A partial reversion of the OHC dysfunction was obtained, with detectable DPOAEs in pejvakin cDNA-treated cochleas (threshold 54.0 ± 10.7 dB), but not in contralateral, untreated cochleas (Figure 7E). DPOAE thresholds were linearly correlated (r^2 = 0.74, p < 0.001) with the number of EGFP-tagged OHCs (Figure 7F), suggesting that the normalization of DPOAE thresholds may be possible if all OHCs could be transduced. The latency of the ABR wave I in response to a 105 dB SPL stimulation decreased significantly (1.38 ± 0.11 ms for the treated ears; n = 6, versus 1.53 ± 0.10 ms
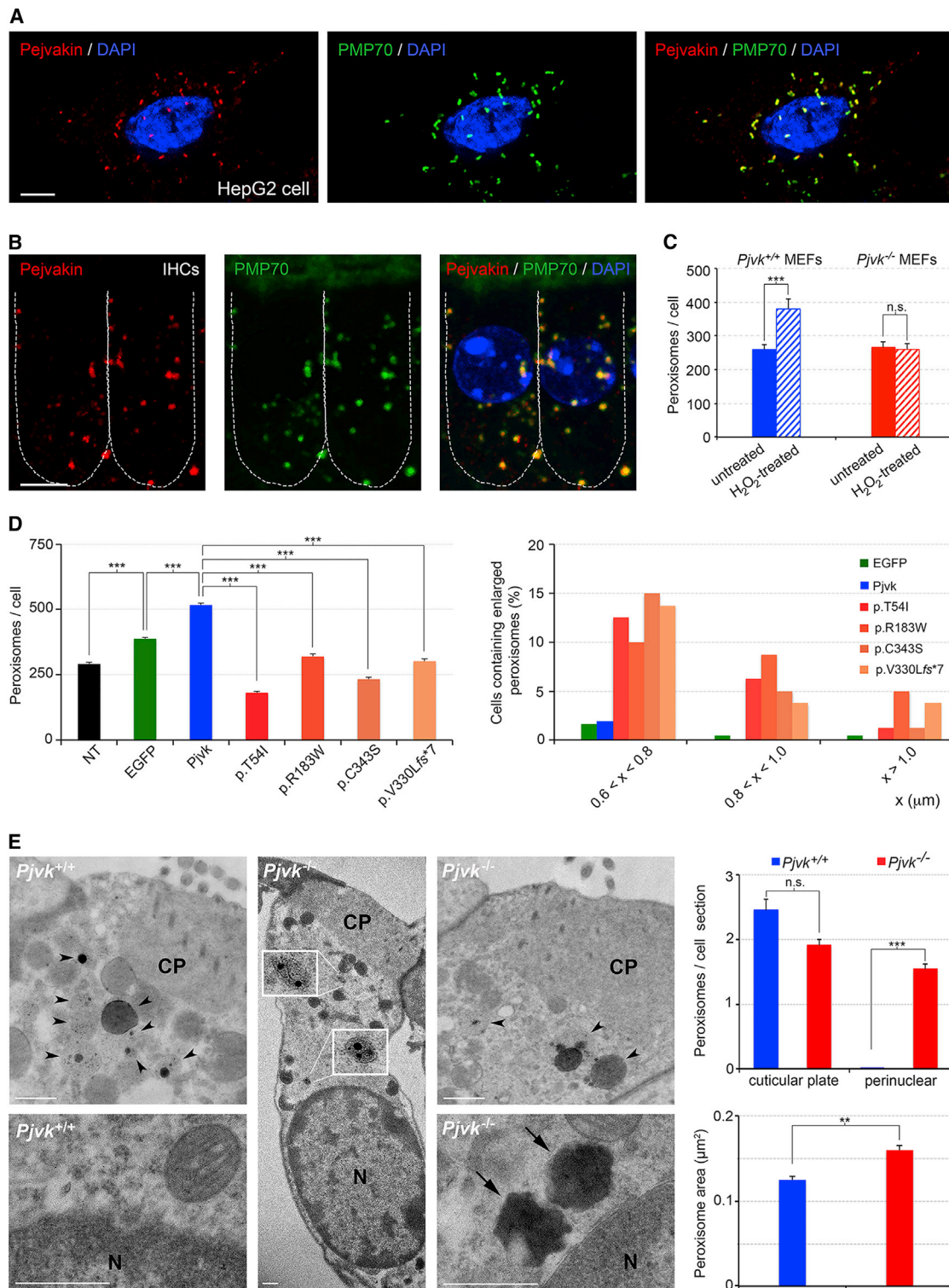
Figure 5. Pejvakin Is a Peroxisome-Associated Protein Involved in the Oxidative Stress-Induced Peroxisomal Proliferation

(A and B) Immunolabeling of PMP70 and endogenous pejvakin in a HepG2 cell (A) and in two P20 $Pjvk^{+/+}$ IHCs (B). See also Figure S6B.

(C) Number of peroxisomes in $Pjvk^{+/+}$ and $Pjvk^{-/-}$ mouse embryonic fibroblasts (MEFs) subjected to 0.5 mM H_2O_2 versus untreated MEFs (n = 30 cells for each condition). See also Figure S7A.

(D) Untransfected HeLa cells (NT) and transfected cells producing either EGFP alone or EGFP, together with the wild-type pejvakin (Pjvk) or a mutated Pjvk (p.T54I, p.R183W, p.C343S, or p.V330Lfs*7). Left panel: bar chart showing the numbers of peroxisome per cell 48 hr after transfection. There were on average

(legend continued on next page)

Figure 6. Effect of Exposure to Loud Sounds on the Cochlear Expression of *Pjvk* and the Number of Peroxisomes in Cochlear Hair Cells and Ganglion Neurons

(A) *Pjvk*, *c-Dct*, *CypA*, *Mpv17*, and *Gpx2* transcript levels assessed by qRT-PCR in P21 *Pjvk*$^{+/+}$ organ of Corti 1, 3, 6, and 18 hr after sound exposure (5–20 kHz, 105 dB SPL for 1 hr). The levels of *c-Fos* and *Hsp70* transcripts were used as positive controls. See also Figure S4B.

(B) Peroxisome proliferation in P21 *Pjvk*$^{+/+}$ hair cells and cochlear ganglion neurons after sound exposure (same conditions as in A). Peroxisomes were counted 48 hr after sound exposure. OHCs, IHCs, and neuronal processes stained for F-actin, myosin VI, and neurofilament protein NF200, respectively. In OHCs and IHCs, the peroxisomes are located below the CP and throughout the cytoplasm, respectively. For OHCs, both a lateral view and a transverse optical section at the level of CP (scheme on the right) are shown. The number of peroxisomes was increased in OHCs, IHCs, and dendrites after sound exposure. N, cell nucleus. ***p < 0.001. Error bars represent the SEM. Scale bars are 5 μm.

See also Figure S4 and Table S1.

Before sound exposure, the numbers of peroxisomes in IHCs of P21 *Pjvk*$^{-/-}$ and AAV2/8-Pjvk-IRES-EGFP-injected *Pjvk*$^{-/-}$ mice did not differ from that of *Pjvk*$^{+/+}$ mice (30.5 ± 1.9, 32.3 ± 2.1, and 36.8 ± 3.0 peroxisomes, mean ± SEM per IHC, n = 60 cells from four *Pjvk*$^{-/-}$ and four AAV2/8-Pjvk *Pjvk*$^{-/-}$ mice, and n = 90 cells from six *Pjvk*$^{+/+}$ mice, respectively; t test, p = 0.11 and p = 0.30, respectively). By contrast, 48 hr after sound exposure (5–20 kHz) at 105 dB SPL for 1 hr, the number of peroxisomes had decreased by 63% in *Pjvk*$^{-/-}$ IHCs (30.5 ± 1.9 and 11.2 ± 1.3 peroxisomes per IHC, n = 75 cells from five unexposed and five sound-exposed *Pjvk*$^{-/-}$ mice, respectively; t test, p < 0.0001), and enlarged PMP70-labeled structures were present close to the nucleus (Figure 7J). In response to the same sound but of a lower intensity, i.e., 97 dB SPL for 1 hr, the number of peroxisomes was unchanged in *Pjvk*$^{-/-}$ IHCs (30.5 ± 1.9 and 34.6 ± 2.3 peroxisomes per IHC, n = 60 cells from four unexposed and four sound-exposed *Pjvk*$^{-/-}$ mice,

for the contralateral, untreated ears; paired t test, p = 0.03) (Figure 7G), and its amplitude increased into the normal range (7.34 ± 0.80 μV versus 2.93 ± 0.92 μV; paired t test, p < 0.001) (Figure 7H), in relation to the number of EGFP-tagged IHCs (r^2 = 0.89 for wave I amplitude, p < 0.001; Figure 7I). No correction of the interwave I-IV latency was observed, as expected (data not shown).

Finally, we investigated the effect of the transduction of *Pjvk*$^{-/-}$ IHCs by AAV2/8-Pjvk-IRES-EGFP on their peroxisomes.

33% more peroxisomes in cells producing both EGFP and Pjvk (n = 200) than in cells producing EGFP alone (n = 150). Right panel: for every range of enlarged peroxisome size, x (0.6–0.8 μm, 0.8–1.0 μm, and >1.0 μm), in two perpendicular directions, the proportion of cells containing at least one peroxisome. See also Figure S7B.

(E) Abnormalities in shape and distribution of peroxisomes in mature *Pjvk*$^{-/-}$ OHCs detected by TEM (P30 *Pjvk*$^{-/-}$ [middle and right] and *Pjvk*$^{+/+}$ [left] OHCs). Insets [middle panel]: enlarged views of individual peroxisomes. In *Pjvk*$^{+/+}$ OHCs, peroxisomes are grouped just under the cuticular plate (CP) (arrowheads), with none detected in the perinuclear region (n = 33 sections, upper bar chart). In *Pjvk*$^{-/-}$ OHCs, some peroxisomes remain under the CP (arrowheads), but catalase-containing structures, misshapen peroxisomes (arrows), are detected in the perinuclear region (n = 24 sections, upper bar chart). Peroxisomes located under the CP are larger in *Pjvk*$^{-/-}$ OHCs (n = 92 peroxisomes) than in *Pjvk*$^{+/+}$ OHCs (n = 89 peroxisomes) (lower bar chart). N, cell nucleus. **p < 0.01, ***p < 0.001. Error bars represent the SEM. Scale bars are 5 μm in (A) and (B) and 0.5 μm in (E).
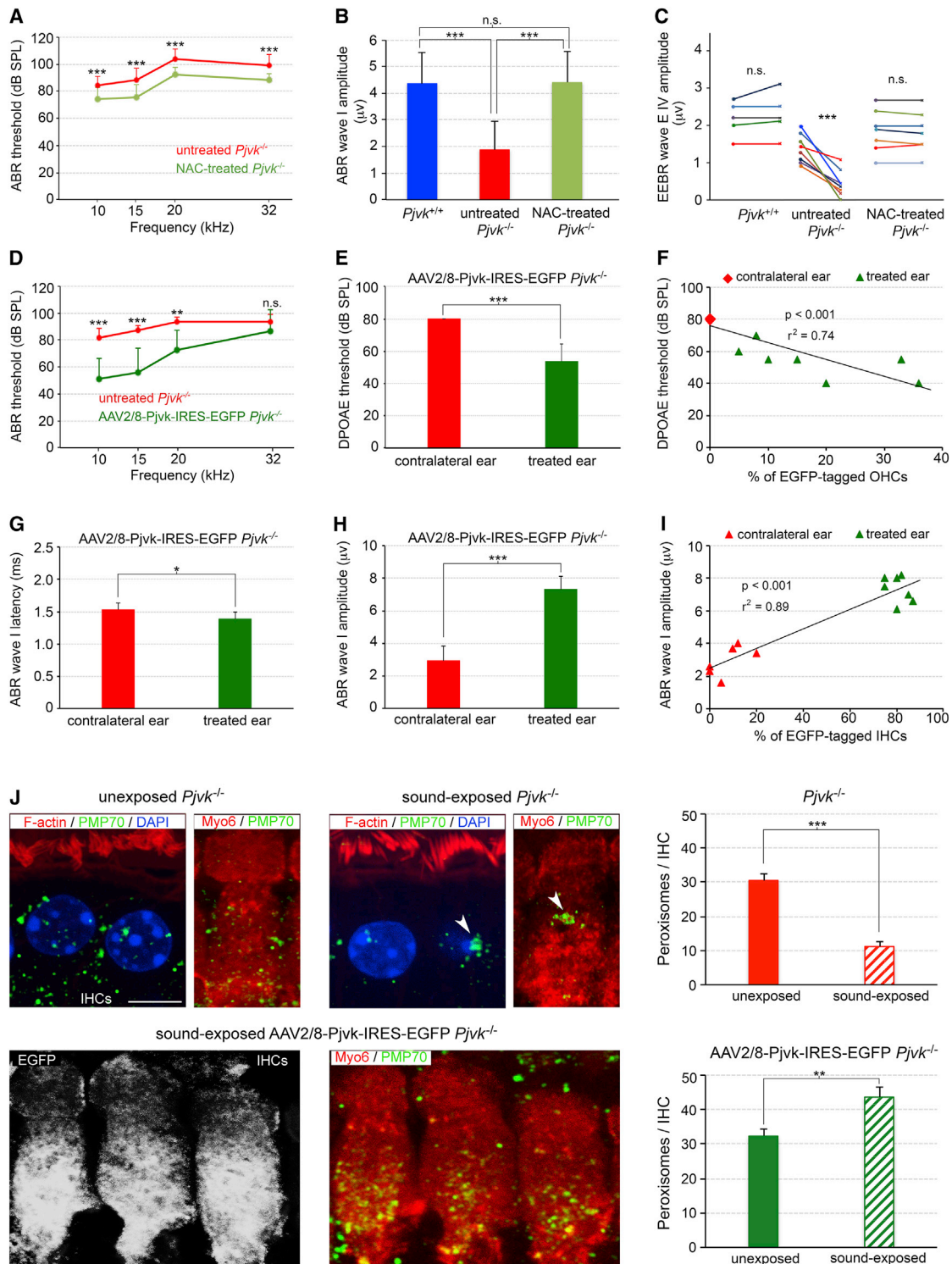
See also Figures S6 and S7.

Figure 7. Therapeutic Approaches in *Pjvk*^{−/−} Mice

(A–C) Effect of N-acetyl cysteine (NAC) on auditory function in *Pjvk*^{−/−} mice. (A) ABR thresholds in untreated versus NAC-treated P21 *Pjvk*^{−/−} mice. (B) ABR wave I amplitude for 10 kHz tone bursts in *Pjvk*^{+/+} and untreated *Pjvk*^{−/−} versus NAC-treated *Pjvk*^{−/−} mice at P21. (C) EEBR wave E IV amplitude before (dots) and after (crosses) controlled electrical stimulation of the cochlear nerve at 200 impulses/s for 1 min in *Pjvk*^{+/+}, untreated *Pjvk*^{−/−}, and NAC-treated *Pjvk*^{−/−} mice.

(D–I) Effect of AAV2/8-Pjvk-IRES-EGFP transferred into the cochlear hair cells on the auditory function of *Pjvk*^{−/−} mice. See also Figure S2B. (D) ABR thresholds at 10, 15, and 20 kHz in AAV2/8-Pjvk-IRES-EGFP-treated versus untreated *Pjvk*^{−/−} mice. (E and H) DPOAE threshold (E) and ABR wave I amplitude (H) at 10 kHz in

(legend continued on next page)

respectively; t test, p = 0.17), and no enlarged PMP70-stained structures were detected (data not shown). The absence of pejvakin thus resulted in defective sound-induced peroxisomal proliferation (both at 105 dB SPL and 97 dB SPL) and, even, in peroxisome degeneration (at 105 dB SPL) in IHCs. In $Pjvk^{-/-}$ mice injected with AAV2/8-Pjvk-IRES-EGFP on P3 and exposed to 105 dB SPL for 1 hr on P21, enlarged PMP70-labeled structures were no longer detected in transduced IHCs, and the number of peroxisomes increased by 35% (32.3 ± 2.1 and 43.7 ± 3.0 peroxisomes per IHC, n = 60 cells from unexposed and exposed transduced $Pjvk^{-/-}$ IHCs, respectively; t test, p = 0.002) (Figure 7J). We conclude that pejvakin re-expression fully protects $Pjvk^{-/-}$ IHCs from the degenerescence of peroxisomes and partially restores their impaired adaptive proliferation.

DISCUSSION

Noise-induced hearing loss (NIHL) is the second most common form of sensorineural hearing impairment after presbycusis in the United States (Dobie, 2008). Here, we describe a genetic form of NIHL, by showing that pejvakin deficiency in mice and DFNB59 patients leads to hypervulnerability to sound, due to a peroxisomal deficiency. To our knowledge, a peroxisomal cause of an isolated (non-syndromic) form of inherited deafness has not been reported yet. The peroxisome emerges as a key organelle in the redox homeostasis of the auditory system, for coping with the overproduction of ROS induced by high levels of acoustic energy.

Acoustic energy is the main determinant of NIHL. The $L_{EX,8\ hr}$ (for level of exposure over an 8-hr workshift) index has been defined such that an $L_{EX,8\ hr}$ of X dB delivers the same energy as a stable sound of X dB played over a period of 8 hr. Chronic occupational exposures to less than 85 dB (or 80 dB, depending on the country) are deemed safe. In $Pjvk^{-/-}$ mice, a single exposure to 63 dB $L_{EX,8\ hr}$ increased hearing thresholds by 30 dB, with full recovery occurring after about 2 weeks. By contrast, a ten times more energetic exposure to a $L_{EX,8\ hr}$ of 73 dB in wild-type mice of the same strain produces only an 18 dB shift in threshold, with a recovery time of 12 hr (Housley et al., 2013). This hypersensitivity of $Pjvk^{-/-}$ mice to noise suggests that the $L_{EX,8\ hr}$ of about 83 dB for a cage of ten pups is sufficient to account for permanent hearing loss in these $Pjvk^{-/-}$ pups, while some of those housed in small numbers in quiet rooms can display near-normal hearing thresholds (see Figure 1C). Likewise, the auditory function of DFNB59 patients was transiently affected by a 57 dB $L_{EX,8\ hr}$ exposure, routinely used in ABR tests.

NIHL involves the excessive production of ROS, overwhelming the antioxidant defense system and causing irreversible oxidative damage to DNA, proteins, and lipids within the cell (Henderson et al., 2006). Noise-induced oxidative stress results in the production of H_2O_2 and other ROS as by-products, thought to derive from the intense solicitation of mitochondrial activity, and several

mouse mutants with mitochondrial defects are prone to NIHL (Oh-lemiller et al., 1999; Brown et al., 2014). Our studies of pejvakin-deficient mouse mutants and rescue experiments targeting the hair cells and auditory neurons unambiguously show that IHCs, OHCs, primary auditory neurons, and neurons of the cochlear nucleus are hypervulnerable to sound in the absence of pejvakin, which is consistent with previous results showing that hair cells and neurons of the auditory system are targets of NIHL (Wang et al., 2002; Kujawa and Liberman, 2009; Imig and Durham, 2005). However, our study goes one step further by implicating a possible common mechanism: peroxisomal failure, the importance of which is demonstrated by the impairment of the redox homeostasis caused by pejvakin deficiency. It also reveals a major cause of the unusually high level of phenotypic variability observed in pejvakin-deficient mice and humans: the difference in sound exposure and the inability of the peroxisomes to cope with the resultant activity-dependent oxidative stress in the absence of pejvakin. Incidentally, this can account for the apparent paradox that mice carrying the R183W mutation in pejvakin displayed a much more severe neural pathway defect than the $Pjvk^{-/-}$ mice (Delmaghani et al., 2006). Due to the preservation of hair cell functions, the auditory neurons of R183W mutant mice should be strongly stimulated, whereas the early permanent damage to cochlear hair cells in $Pjvk^{-/-}$ mice acts as a protective "muffler" of the neuronal pathway.

In mammals, the number and metabolic functions of peroxisomes differ between cell types. However, all cell types are able to adapt rapidly to modifications in physiological conditions by changing the number, shape, size, and molecular content of peroxisomes, resulting in considerable functional plasticity of these organelles (Schrader et al., 2012; Smith and Aitchison, 2013). Our experiments on $Pjvk^{-/-}$ and $Pjvk^{+/+}$ mouse embryonic fibroblasts stressed with H_2O_2 showed that pejvakin is critically involved in the oxidative stress-induced proliferation of peroxisomes through growth and fission of pre-existing peroxisomes. The molecular machinery underlying this adaptive process is still poorly understood beyond the involvement of Pex11α (Li et al., 2002). Of note, the absence of pejvakin only affects the proliferation of peroxisomes from pre-existing peroxisomes, but not the constitutive biogenesis of this organelle. Accordingly, structural abnormalities of peroxisomes in $Pjvk^{-/-}$ mice became apparent only after hearing onset, in the context of the oxidative stress produced by noise exposure. By contrast, the PEX gene defects causing Zellweger syndrome spectrum (ZSS) disorders (Water-ham and Ebberink, 2012) and rhizomelic chondrodysplasia punctata affect the constitutive biogenesis of peroxisomes. Hearing impairment in ZSS disorders involves a severe impairment of neuronal conduction and has been attributed to defects in the synthesis of two essential myelin sheath components—plasmalogens and docosahexaenoic acid—which is critically dependent on peroxisomes. Our results suggest that ZSS also

---

treated versus untreated contralateral ears. (G) ABR wave I latency in treated versus untreated contralateral ears. (F) Correlation between DPOAE thresholds and the proportion of EGFP-tagged (i.e., transduced) OHCs. Six untreated ears have no recordable DPOAE (threshold arbitrarily set at 80 dB SPL; red diamond). (I) Correlation between ABR wave I amplitude at 10 kHz, 105 dB SPL, and the percentage of transduced IHCs (EGFP tagged).
(J) Effect of AAV2/8-Pjvk-IRES-EGFP on the peroxisomes in $Pjvk^{-/-}$ IHCs. Upper and lower panels show and quantify (bar charts) the peroxisomes in untreated mice 48 hr after sound exposure (5–20 kHz, 105 dB SPL for 1 hr) (peroxisome abnormalities are indicated by arrowheads). Error bars represent the SD in (A–I) and the SEM in (J). ns, not significant; *p < 0.05, **p < 0.01, ***p < 0.001.

includes a defective redox balance in the hair cells and neurons of the auditory system.

In the context of noise exposure, the upregulation of *Pjvk* transcription in the cochlea and the subsequent peroxisome proliferation in the hair cells and auditory neurons of wild-type mice suggest that pejvakin-dependent peroxisome proliferation in the auditory system is part of the physiological response to high levels of acoustic energy that result in increased amounts of ROS. This and the marked oxidative stress detected in the *Pjvk*$^{-/-}$ cochlea imply that the proliferation of peroxisomes plays an antioxidant role, similar to that reported in other cell types (Santos et al., 2005; Diano et al., 2011). The rapid elevation of the hearing threshold in *Pjvk*$^{-/-}$ mice in response to low-energy sounds and the increase in interwave I-IV latency observed in DFNB59 patients within a few seconds are consistent with an activity-dependent H_2O_2 production that, due to impaired cellular redox homeostasis, results in concentrations of H_2O_2 high enough to impact on the activity of various target proteins including ion channels and transporters (Rice, 2011). The worsening of hearing sensitivity, 2 days later, in the mutant mice lacking pejvakin, exacerbated by putting back the mice in a noisy environment, fits the picture of the absence of sound-induced biogenesis of peroxisomes (with their degeneration occurring in a high acoustic energy environment). We thus conclude that the hypervulnerability of *Pjvk*$^{-/-}$ mice and DFNB59 patients to sound does not result simply from an exacerbation, by sound, of a pre-existing redox-balance defect, but is the consequence of impaired adaptive proliferation of peroxisomes in the absence of pejvakin. Both defective peroxisome proliferation in IHCs of *Pjvk*$^{-/-}$ mice in response to sound exposure and its partial recovery by pejvakin cDNA transfer support this conclusion. A full recovery of the adaptive peroxisome proliferation produced by sound exposure may require higher concentrations of pejvakin or the sound-induced modulation of *Pjvk* transcription (see Figure 6A), which was missing in our rescue experiments (pejvakin cDNA expression being driven by a constitutive promoter).

In patients with hearing impairment, the amplification of sound by hearing aids or direct electrical stimulation of the auditory nerve by a cochlear implant delivers a stimulus with an energy level similar to that shown here to worsen the hearing impairment of *Pjvk*$^{-/-}$ mice within 1 min of sound exposure. Therefore, in cases of peroxisomal deficiency, as in DFNB59, specific protection against redox homeostasis failure is essential. Patients with such conditions should avoid noisy environments and a beneficial effect of hearing devices should require an antioxidant protection. N-acetyl cysteine was the only antioxidant drug tested here to display some, albeit limited, efficacy. By contrast, AAV-mediated gene therapy could potentially provide full protection. Finally, deciphering the sound-stress-induced protective signaling pathway involving pejvakin might lead to the discovery of therapeutic agents for NIHL.

EXPERIMENTAL PROCEDURES

Audiological Studies in Mice

Auditory tests were performed in an anechoic room, on anesthetized animals whose core temperatures were maintained at 37°C (see the Supplemental Experimental Procedures).

Audiological Tests in Patients

Informed consent was obtained from all the subjects included in the study. Pure-tone audiometry was performed with air- and bone-transmitted tones. Hearing impairment was assessed objectively, by measuring ABRs and transient-evoked otoacoustic emissions (TEOAEs). The nonlinear TEOAE recording procedure was used (derived from the ILO88 system), making it possible to extract TEOAEs from linear reflection artifacts from the middle ear, and to evaluate background noise. TEOAE responses were analyzed in 1-kHz-wide bands centered on 1, 2, 3 and 4 kHz.

Generation of an Anti-pejvakin Monoclonal Antibody

The 3′ end of the coding sequence of the Pjvk cDNA (NCBI:NM_001080711.2) was inserted into a pGST-parallel-2 vector (derived from pGEX-4T-1; Amersham). The resultant construct, encoding the C-terminal region of pejvakin (residues 290–352; RefSeq:NP_001074180.1) fused to an N-terminal glutathione S-transferase tag, was introduced into *Escherichia coli* BL21-Gold (DE3)-competent cells (Stratagene). The pejvakin protein fragment was purified on a glutathione-Sepharose 4B column, then subjected to size-exclusion chromatography and used as the antigen for immunization. Antibodies were produced by immunizing *Pjvk*$^{-/-}$ mice. An immunoglobulin G monoclonal antibody (K_D of 6×10^{-8} M), Pjvk-G21, was selected by ELISA on immunogen-coated plates.

Statistical Analyses

Quantitative data are presented as mean ± SD, unless otherwise mentioned. Statistical analyses were performed using GraphPad. Data were analyzed by paired or unpaired Student's t tests and, for multiple comparisons, either by one-way or two-way ANOVA or by t tests with the Bonferroni correction. Statistical significance of the differences observed between groups is defined as $p < 0.05$.

ACCESSION NUMBER

The accession number for the transcriptomic data reported in this paper is deposited in GEO:GSE72722.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.023.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

REFERENCES

Ashmore, J. (2008). Cochlear outer hair cell motility. Physiol. Rev. *88*, 173–210.

Avan, P., Büki, B., and Petit, C. (2013). Auditory distortions: origins and functions. Physiol. Rev. *93*, 1563–1619.

Borck, G., Rainshtein, L., Hellman-Aharony, S., Volk, A.E., Friedrich, K., Taub, E., Magal, N., Kanaan, M., Kubisch, C., Shohat, M., and Basel-Vanagaite, L. (2012). High frequency of autosomal-recessive DFNB59 hearing loss in an isolated Arab population in Israel. Clin. Genet. *82*, 271–276.

Brown, K.D., Maqsood, S., Huang, J.Y., Pan, Y., Harkcom, W., Li, W., Sauve, A., Verdin, E., and Jaffrey, S.R. (2014). Activation of SIRT3 by the NAD$^+$ precursor nicotinamide riboside protects from noise-induced hearing loss. Cell Metab. *20*, 1059–1068.

Collin, R.W., Kalay, E., Oostrik, J., Caylan, R., Wollnik, B., Arslan, S., den Hollander, A.I., Birinci, Y., Lichtner, P., Strom, T.M., et al. (2007). Involvement of DFNB59 mutations in autosomal recessive nonsyndromic hearing impairment. Hum. Mutat. *28*, 718–723.

Delmaghani, S., del Castillo, F.J., Michel, V., Leibovici, M., Aghaie, A., Ron, U., Van Laer, L., Ben-Tal, N., Van Camp, G., Weil, D., et al. (2006). Mutations in the gene encoding pejvakin, a newly identified protein of the afferent auditory pathway, cause DFNB59 auditory neuropathy. Nat. Genet. *38*, 770–778.

Diano, S., Liu, Z.W., Jeong, J.K., Dietrich, M.O., Ruan, H.B., Kim, E., Suyama, S., Kelly, K., Gyengesi, E., Arbiser, J.L., et al. (2011). Peroxisome proliferation-associated control of reactive oxygen species sets melanocortin tone and feeding in diet-induced obesity. Nat. Med. *17*, 1121–1127.

Dobie, R.A. (2008). The burdens of age-related and occupational noise-induced hearing loss in the United States. Ear Hear. *29*, 565–577.

Ebberink, M.S., Koster, J., Visser, G., Spronsen, Fv., Stolte-Dijkstra, I., Smit, G.P., Fock, J.M., Kemp, S., Wanders, R.J., and Waterham, H.R. (2012). A novel defect of peroxisome division due to a homozygous non-sense mutation in the PEX11β gene. J. Med. Genet. *49*, 307–313.

Ebermann, I., Walger, M., Scholl, H.P., Charbel Issa, P., Lüke, C., Nürnberg, G., Lang-Roth, R., Becker, C., Nürnberg, P., and Bolz, H.J. (2007). Truncating mutation of the DFNB59 gene causes cochlear hearing impairment and central vestibular dysfunction. Hum. Mutat. *28*, 571–577.

Henderson, D., Bielefeld, E.C., Harris, K.C., and Hu, B.H. (2006). The role of oxidative stress in noise-induced hearing loss. Ear Hear. *27*, 1–19.

Housley, G.D., Morton-Jones, R., Vlajkovic, S.M., Telang, R.S., Paramananthasivam, V., Tadros, S.F., Wong, A.C., Froud, K.E., Cederholm, J.M., Sivakumaran, Y., et al. (2013). ATP-gated ion channels mediate adaptation to elevated sound levels. Proc. Natl. Acad. Sci. USA *110*, 7494–7499.

Imig, T.J., and Durham, D. (2005). Effect of unilateral noise exposure on the tonotopic distribution of spontaneous activity in the cochlear nucleus and inferior colliculus in the cortically intact and decorticate rat. J. Comp. Neurol. *490*, 391–413.

Kujawa, S.G., and Liberman, M.C. (2009). Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. J. Neurosci. *29*, 14077–14085.

Li, X., Baumgart, E., Dong, G.X., Morrell, J.C., Jimenez-Sanchez, G., Valle, D., Smith, K.D., and Gould, S.J. (2002). PEX11alpha is required for peroxisome proliferation in response to 4-phenylbutyrate but is dispensable for peroxisome proliferator-activated receptor alpha-mediated peroxisome proliferation. Mol. Cell. Biol. *22*, 8226–8240.

Lopez-Huertas, E., Charlton, W.L., Johnson, B., Graham, I.A., and Baker, A. (2000). Stress induces peroxisome biogenesis genes. EMBO J. *19*, 6770–6777.

Mizuno, Y., Kurochkin, I.V., Herberth, M., Okazaki, Y., and Schönbach, C. (2008). Predicted mouse peroxisome-targeted proteins and their actual subcellular locations. BMC Bioinformatics *9* (*Suppl 12*), S16.

Møller, A.R., and Jannetta, P.J. (1983). Interpretation of brainstem auditory evoked potentials: results from intracranial recordings in humans. Scand. Audiol. *12*, 125–133.

Ohlemiller, K.K., Wright, J.S., and Dugan, L.L. (1999). Early elevation of cochlear reactive oxygen species following noise exposure. Audiol. Neurootol. *4*, 229–236.

Oliver, D., Taberner, A.M., Thurm, H., Sausbier, M., Arntz, C., Ruth, P., Fakler, B., and Liberman, M.C. (2006). The role of BKCa channels in electrical signal encoding in the mammalian auditory periphery. J. Neurosci. *26*, 6181–6189.

Rice, M.E. (2011). H_2O_2: a dynamic neuromodulator. Neuroscientist *17*, 389–406.

Robles, L., and Ruggero, M.A. (2001). Mechanics of the mammalian cochlea. Physiol. Rev. *81*, 1305–1352.

Santos, M.J., Quintanilla, R.A., Toro, A., Grandy, R., Dinamarca, M.C., Godoy, J.A., and Inestrosa, N.C. (2005). Peroxisomal proliferation protects from beta-amyloid neurodegeneration. J. Biol. Chem. *280*, 41057–41068.

Schrader, M., and Fahimi, H.D. (2006). Peroxisomes and oxidative stress. Biochim. Biophys. Acta *1763*, 1755–1766.

Schrader, M., Bonekamp, N.A., and Islinger, M. (2012). Fission and proliferation of peroxisomes. Biochim. Biophys. Acta *1822*, 1343–1357.

Schwander, M., Sczaniecka, A., Grillet, N., Bailey, J.S., Avenarius, M., Najmabadi, H., Steffy, B.M., Federe, G.C., Lagler, E.A., Banan, R., et al. (2007). A forward genetics screen in mice identifies recessive deafness traits and reveals that pejvakin is essential for outer hair cell function. J. Neurosci. *27*, 2163–2175.

Smith, J.J., and Aitchison, J.D. (2013). Peroxisomes take shape. Nat. Rev. Mol. Cell Biol. *14*, 803–817.

Starr, A., and Rance, G. (2015). Auditory neuropathy. Handb. Clin. Neurol. *129*, 495–508.

Tang, X.D., Garcia, M.L., Heinemann, S.H., and Hoshi, T. (2004). Reactive oxygen species impair Slo1 BK channel function by altering cysteine-mediated calcium sensing. Nat. Struct. Mol. Biol. *11*, 171–178.

Wang, Y., Hirose, K., and Liberman, M.C. (2002). Dynamics of noise-induced cellular injury and repair in the mouse cochlea. J. Assoc. Res. Otolaryngol. *3*, 248–268.

Waterham, H.R., and Ebberink, M.S. (2012). Genetics and molecular basis of human peroxisome biogenesis disorders. Biochim. Biophys. Acta *1822*, 1430–1441.

Cell

Vesicle-Mediated Steroid Hormone Secretion in *Drosophila melanogaster*

Graphical Abstract
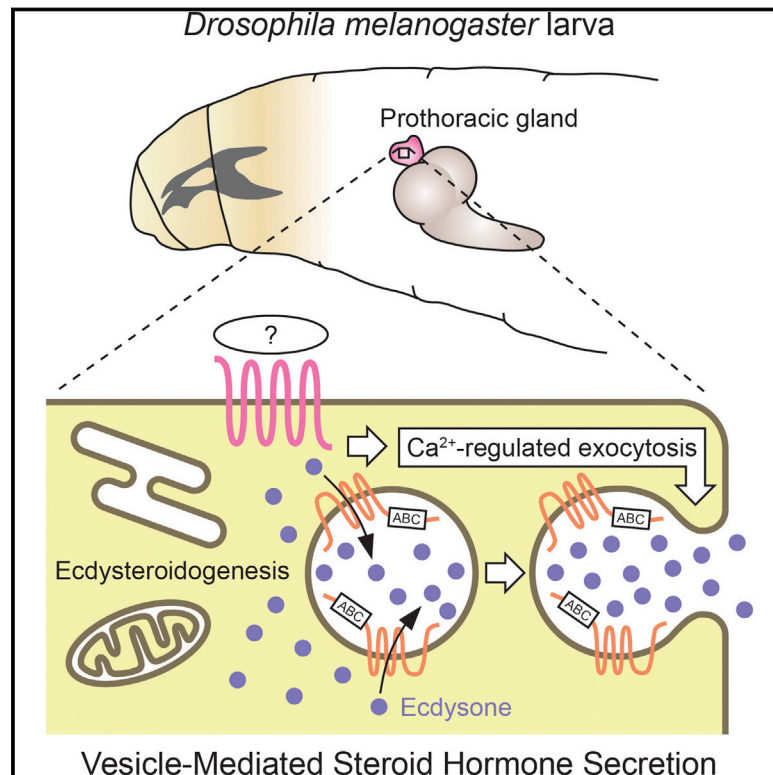
Authors

Naoki Yamanaka, Guillermo Marqués, Michael B. O'Connor

Correspondence

naoki.yamanaka@ucr.edu (N.Y.), moconnor@umn.edu (M.B.O.)

In Brief

The release of the steroid hormone ecdysone in *Drosophila* is mediated through a vesicular trafficking mechanism regulated via calcium signaling, suggesting that steroid hormones may be actively released from cells, as opposed to simply diffusing across the membrane.

Highlights

- Conventional "free diffusion" model of steroid hormone release should be revisited

- Components for secretory vesicle exocytosis are required for ecdysone secretion

- Calcium-regulated vesicle exocytosis is coupled with ecdysone release

- An ABC transporter is required to fill secretory vesicles with ecdysone

CrossMark

CellPress

Vesicle-Mediated Steroid Hormone Secretion in *Drosophila melanogaster*

Naoki Yamanaka,[1,2,*] Guillermo Marqués,[3] and Michael B. O'Connor[1,*]

[1]Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN 55455, USA
[2]Department of Entomology, Institute for Integrative Genome Biology, Center for Disease Vector Research, University of California, Riverside, Riverside, CA 92521, USA
[3]University Imaging Centers, University of Minnesota, Minneapolis, MN 55455, USA
*Correspondence: naoki.yamanaka@ucr.edu (N.Y.), moconnor@umn.edu (M.B.O.)
http://dx.doi.org/10.1016/j.cell.2015.10.022

SUMMARY

Steroid hormones are a large family of cholesterol derivatives regulating development and physiology in both the animal and plant kingdoms, but little is known concerning mechanisms of their secretion from steroidogenic tissues. Here, we present evidence that in *Drosophila*, endocrine release of the steroid hormone ecdysone is mediated through a regulated vesicular trafficking mechanism. Inhibition of calcium signaling in the steroidogenic prothoracic gland results in the accumulation of unreleased ecdysone, and the knockdown of calcium-mediated vesicle exocytosis components in the gland caused developmental defects due to deficiency of ecdysone. Accumulation of synaptotagmin-labeled vesicles in the gland is observed when calcium signaling is disrupted, and these vesicles contain an ABC transporter that functions as an ecdysone pump to fill vesicles. We propose that trafficking of steroid hormones out of endocrine cells is not always through a simple diffusion mechanism as presently thought, but instead can involve a regulated vesicle-mediated release process.

INTRODUCTION

Steroid hormones are an important class of bioactive molecules in both animal and plant kingdoms that regulate a wide variety of physiological processes including immune response, salt and water balance, glucose metabolism, and sexual maturation during juvenile development (Sapolsky et al., 2000; Sisk and Foster, 2004). In insect larvae, the primary precursor steroid hormone ecdysone (E) is produced in the prothoracic gland (PG). After its release into the circulatory system, E is taken up in peripheral tissues such as the gut and fat body where it is converted to 20-hydroxyecdysone (20E). This is the active derivative that regulates larval molt timing and the onset of metamorphosis leading to the formation of sexually mature adults (Yamanaka et al., 2013a).

The biosynthetic pathways of steroid hormone production have been extensively studied in diverse animal species, and many of the key enzymes have been identified and well characterized (Ghayee and Auchus, 2007; Huang et al., 2008; Miller, 2013). In insects, E biosynthesis is stimulated by extracellular signaling molecules such as the prothoracicotropic hormone (PTTH), which binds to its receptor Torso to induce the expression of genes encoding steroidogenic enzymes (Rewitz et al., 2009; Yamanaka et al., 2013a).

In contrast to the extensive literature describing studies on steroidogenic processes, very little is known about the mechanisms that regulate release of steroid hormones from endocrine tissues. Indeed, the textbook view is that lipophilic steroid hormones simply enter and exit cells by diffusion across lipid bilayers (Raven and Johnson, 2002; Sherwood, 2011; White and Porterfield, 2012). However, this prevailing assumption has not been extensively tested in vivo, and the limited studies described so far primarily used in vitro or in silico approaches (Oren et al., 2004; Watanabe et al., 1991).

Given the scarcity of knowledge concerning this fundamental aspect of endocrinology, we used molecular genetic tools to investigate the mechanism of E release from the PG in *Drosophila melanogaster*. We found that blocking calcium signaling through RNAi-mediated knockdown of the *inositol 1,4,5,-tris-phosphate receptor* (*IP3R*) in the PG leads to a buildup of E and a decrease of 20E in source and target tissues, respectively. This results in severe delay or larval developmental arrest that can be rescued by feeding larvae E. Identical developmental defects were observed in larvae in which cellular components normally involved in calcium-mediated vesicle exocytosis, such as Rab3, UNC-13, or synaptotagmin 1 (Syt1), were depleted in the PG. Moreover, GCaMP imagining of the PG just prior to metamorphosis revealed spontaneous calcium signaling that was attenuated by RNAi-mediated knockdown of the upstream signaling component that couples G protein-coupled receptors (GPCRs) to calcium release. Furthermore, the accumulation of Syt1-positive vesicles was observed when calcium signaling was blocked in the PG, suggesting that calcium-mediated vesicle exocytosis is required for E release. Consistent with this notion, we identified an ABC transporter found in these Syt1-positive vesicles and show that it transports E across a lipid bilayer in vitro. Taken together, these results support a new hypothesis that transport of steroid hormones across lipid bilayers can involve a regulated vesicle-release process instead of, or in addition to, passive or facilitated diffusion mechanisms.
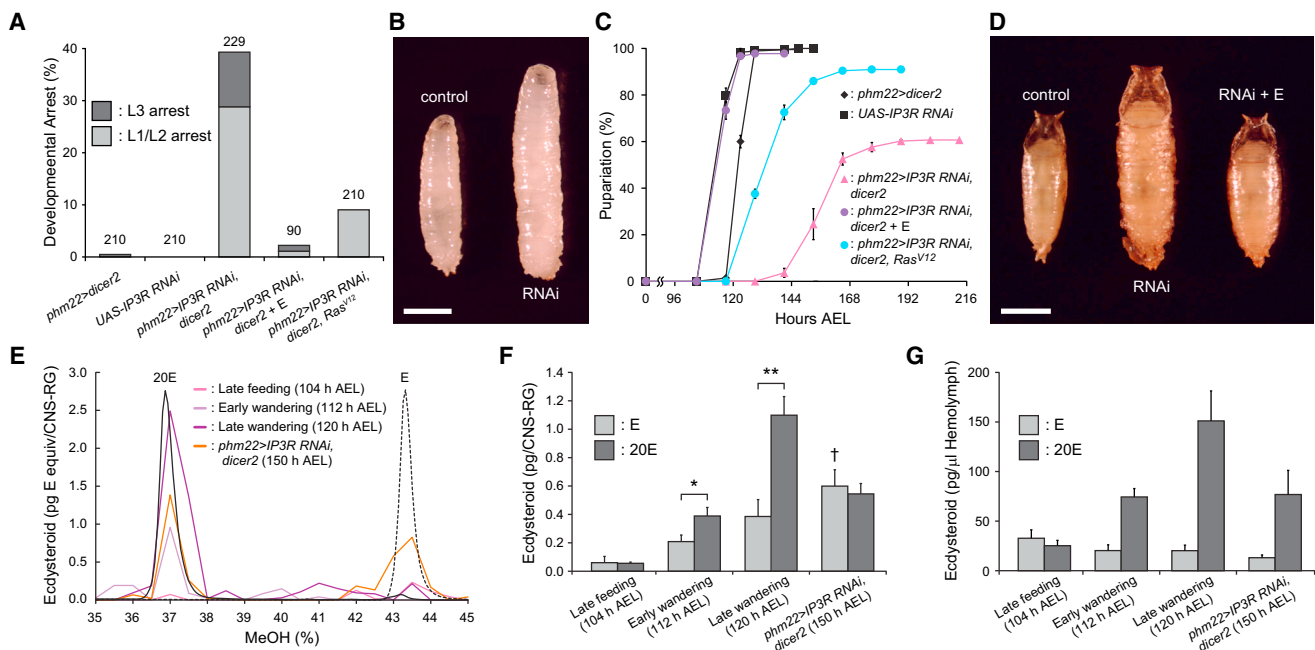
Figure 1. *IP3R* Is Required for E Secretion from the PG

(A) *IP3R*-knockdown in the PG causes polyphasic growth arrest. Percentages of developmentally arrested larvae for each genotype are shown. Percentages of larvae arrested at first or second instar (L1/L2) are indicated in light gray, whereas those arrested at third instar (L3) are shown in dark gray. E feeding is indicated by + E. Numbers of animals tested are shown on top of each bar.

(B) *IP3R*-knockdown in the PG causes overgrowth at L3. Representative image of a wandering control larva (control, *phm22>dicer2*) and an *IP3R*-knockdown larva arrested as L3 (RNAi, *phm22 > IP3R RNAi, dicer2*). Scale bar, 1 mm.

(C) *IP3R*-knockdown in the PG causes developmental delay. Developmental time to pupariation of non-arrested larvae is shown. AEL, after egg laying. Data are represented as mean ± SEM of three to seven independent experiments.

(D) *IP3R*-knockdown in the PG leads to the formation of overgrown pupae. Representative image of a control pupa (control, *phm22>dicer2*), an *IP3R*-knockdown pupa (RNAi, *phm22 > IP3R RNAi, dicer2*) and an *IP3R*-knockdown pupa rescued by E feeding (RNAi + E, *phm22 > IP3R RNAi, dicer2* with E feeding) is shown. Scale bar, 1 mm.

(E) Separation of E and 20E by a methanol (MeOH) gradient on reverse-phase HPLC. UV absorbance profiles at 248 nm of standard E (dashed black line) and 20E (solid black line) are shown on an arbitrary scale. The relative quantity of ecdysteroids in CNS-RG complexes of different groups of animals are shown in solid colored lines as E equivalent (pg) in each fraction. Fractions corresponding to 36%–38% MeOH and 42.5%–44.5% MeOH were pooled and used for 20E and E quantification, respectively.

(F) Quantity of ecdysteroids in CNS-RG complexes at different developmental times and larval genotypes. The E titer is indicated in light gray, whereas that of 20E is shown in dark gray. Each bar represents mean ± SEM of three independent sample preparations. *$p < 0.05$; **$p < 0.01$ from Student's t test. †$p < 0.05$ compared to the amount of E in late feeding control larvae (ANOVA with Tukey's post hoc test).

(G) Quantity of ecdysteroids in the hemolymph at different developmental times and larval genotypes. The E titer is indicated in light gray, whereas that of 20E is shown in dark gray. Each bar represents mean ± SEM of three independent sample preparations.

RESULTS

IP3R-Mediated Calcium Signaling Is Required for E Secretion from the PG

Studies using isolated PGs of lepidopteran insect species have long suggested a key role for calcium in stimulating E production and release in response to PTTH (Huang et al., 2008). To test this in *Drosophila*, we conducted PG-specific knockdown of *IP3R*, which encodes an intracellular calcium-release channel. *IP3R* is highly expressed in the PG and mutants have growth defects due to low systemic levels of E (Venkatesh and Hasan, 1997), but direct links between the IP3R function and E production or release have yet to be tested. When *IP3R* RNAi was induced using the PG driver *phm22-Gal4*, we observed polyphasic growth arrest throughout larval development (Figure 1A). Those larvae that arrested in the third instar stage showed an overgrowth

phenotype due to an extended larval feeding period, which is commonly observed for E-deficient animals (Figure 1B) (Cáceres et al., 2011; Ou et al., 2011; Rewitz et al., 2009; Talamillo et al., 2008). Feeding E to these larvae rescued the arrest phenotype, further suggesting that the systemic E level is low in these animals. Even those individuals that eventually initiated metamorphosis (∼60%) only did so after a significant delay (Figure 1C) and also exhibited an overgrowth phenotype (Figure 1D), both of which were fully rescued by E feeding. Importantly, however, neither the larval arrest nor pupariation delay phenotype was fully rescued by overexpressing Ras^V12, the active form of Ras that is able to completely rescue a PTTH signaling deficiency (Figures 1A and 1C) (Rewitz et al., 2009). Such partial rescue strongly suggests that, unlike previous assumptions based on moth studies (Huang et al., 2008), the intracellular calcium release mediated by IP3R is not solely functioning upstream of the MAPK signaling

cascade and E synthesis, but is likely regulating other processes in the *Drosophila* PG.

One hypothesis we considered was that calcium signaling might play a role in E release from the PG as opposed to, or in addition to, E synthesis. As a first test of this hypothesis, we sought to quantify E and 20E in the PG of *phm22 > IP3R RNAi, dicer2* larvae. The PG is part of a composite endocrine organ called the ring gland (RG), which is attached to the CNS. The small size of the RG presents a substantial dissection challenge. Therefore, we dissected the larger CNS-RG complex, extracted steroids with methanol, and separated E from 20E by high-performance liquid chromatography (HPLC). The E and 20E containing fractions were quantified using an ELISA for ecdysteroids (Figure 1E). This analysis revealed that the titers of both E and 20E increase in the CNS-RG complex of control larvae during the wandering stage as they prepare for metamorphosis (Figure 1F). This observation is consistent with the notion that the E biosynthetic activity of the PG is stimulated by PTTH during this period and the increased levels of 20E in the hemolymph help initiate wandering behavior when taken up by the brain (Yamanaka et al., 2013b). In control larvae, the titer of 20E was substantially higher than that of E in the CNS-RG complex during the wandering stage, indicating that the release of E into the hemolymph and its conversion to 20E happens quickly under normal conditions. Indeed, the 20E titer in the hemolymph increased rapidly in wandering control larvae, while the E titer remained constant or decreased during this stage (Figure 1G). In the case of *phm22 > IP3R RNAi, dicer2* larvae, however, the ecdysteroid composition in the CNS-RG complex was markedly different; the E titer was elevated above that seen at any stage in control larvae, and the titer of 20E was less instead of more than E (Figure 1F). For *IP3R*-knockdown animals that are developmentally delayed, we used wandering as a behavioral trait and used 150 hr after egg laying (AEL) wandering larvae for hormone titer measurement, in order to ensure that we are examining hormone titers at the same developmental stage. Consistent with this notion, the hemolymph levels of ecdysteroids in these *IP3R*-knockdown animals were comparable to those of early wandering control animals (Figure 1G). From these results, we infer that: (1) E is still produced in the PG even when the intracellular calcium release is downregulated by *IP3R RNAi*, and (2) the release of E into the hemolymph is inhibited upon *IP3R* knockdown, resulting in the accumulation of E in the PG.

Components that Regulate Secretory Vesicle Exocytosis Are Required for Proper PG Function

The above results led us to hypothesize that E secretion from the PG is not a fully passive or facilitated diffusion process, but instead may employ a regulatory mechanism that is under the control of calcium signaling. This is reminiscent of secretory vesicle exocytosis in neurons and endocrine cells, where the fusion of vesicles containing neurotransmitters, or various types of hormones, with the plasma membrane is tightly regulated by multiple components that sense intracellular calcium concentrations (Rizo and Rosenmund, 2008; Sudhof, 2004). Therefore, we conducted RNAi knockdown screening of components known to be either upstream regulators of intracellular calcium release or downstream effectors of calcium-regulated secretory vesicle exocytosis (Figure 2A; Table 1).

There are two distinct genes encoding intracellular calcium-release channels in *Drosophila*: *IP3R* and *Ryanodine receptor* (*RyR*) (Sorrentino et al., 2000). The knockdown of *IP3R* in the PG showed the larval arrest phenotype as described above, whereas that of *RyR* did not cause any discernible defect (Table 1). IP3R is activated by inositol 1,4,5,-tris-phosphate produced by phospholipase C (PLC) enzymes, which are encoded by three genes in *Drosophila* (Shortridge and McKay, 1995). PG-specific RNAi knockdown of one of them, *Plc21C*, caused the larval arrest phenotype (Table 1). *Plc21C* encodes a member of PLCβ class of enzymes, which are typically activated by the Gαq subunit of heterotrimeric G proteins coupled with GPCRs. Although RNAi-mediated knockdown of *Gaq* did not cause developmental arrest, we identified *CG30054* and *CG17760* as two adjacent genes that are highly homologous to *Gaq* (Figures S1A and S1B). PG-specific knockdown of *CG30054*, but not that of *CG17760*, caused a developmental defect similar to knockdown of *Plc21C* and *IP3R* (Table 1). Taken together, these results indicate that the calcium mobilization from intracellular stores through IP3R is regulated by GPCR(s) in the PG (Figure 2A), further suggesting that there is an as-yet-unknown signaling pathway acting in parallel with the PTTH/Torso/MAPK pathway to facilitate E release as opposed to its production.

The ternary SNARE complex that consists of synaptobrevin/VAMP, SNAP-25, and syntaxin plays a pivotal role in various types of vesicle exocytosis in eukaryotes ranging from yeast to human (Li and Chin, 2003; Richmond and Broadie, 2002; Rizo and Rosenmund, 2008; Sudhof, 2004). Indeed, the PG-specific knockdown of *Syb*, one of the two synaptobrevins in *Drosophila*, causes the larval arrest phenotype, supporting the hypothesis that calcium-regulated vesicle exocytosis is required for E secretion (Table 1). The SNARE complex, however, is involved in all intracellular membrane fusion events, and each organism has multiple SNARE proteins that are localized to distinct membrane compartments to specify intracellular compartment identity (Li and Chin, 2003). These diverse SNARE complex functions make it difficult to interpret the above result, since E synthesis in the PG involves the trafficking of synthetic intermediates between organelles, and the disruption of this process is expected to cause similar developmental defects. Likewise, the necessity of the exocyst complex components in the PG (Table 1) (Andrews et al., 2002) is difficult to attribute solely to its potential role in E secretion, since this complex is required for multiple membrane trafficking events, including the transport of vesicles carrying transmembrane receptors from the *trans*-Golgi network to the plasma membrane (Langevin et al., 2005; Murthy et al., 2003). That knockdown of exocyst complex members does disrupt trafficking of transmembrane proteins to the plasma membrane is clearly shown by the depletion of mCD8-GFP in the plasma membrane of PG cells of larvae expressing *phm22 > Sec10 RNAi* (Figure S2). Such a defect will likely disrupt several signaling pathways including PTTH and insulin signaling, both of which require the transport of their receptors to the PG plasma membrane for high level E production (Yamanaka et al., 2013a).

In light of these difficulties in interpreting the phenotypes produced by knockdown of general secretory machinery subunits,
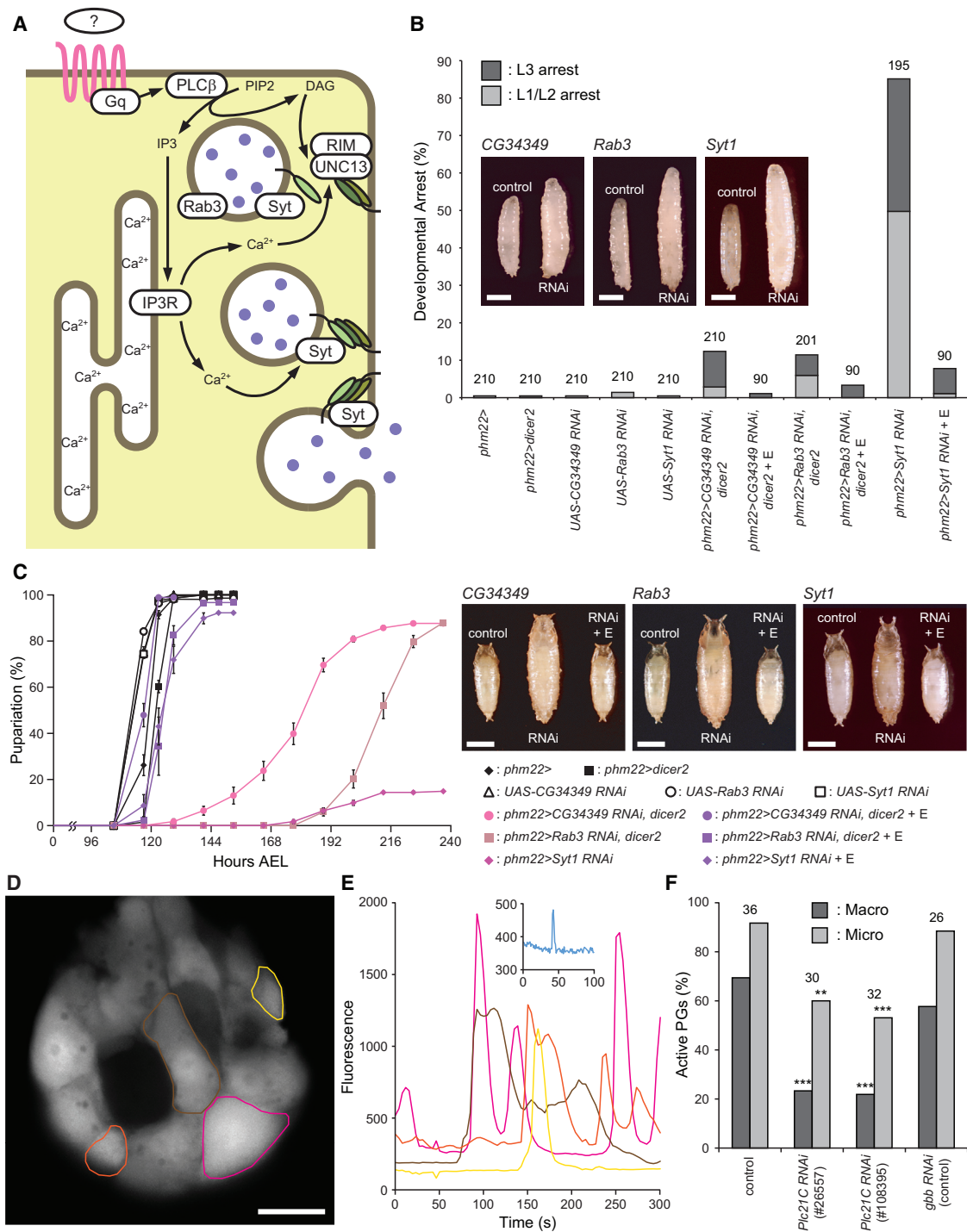
Figure 2. Regulatory Components of Secretory Vesicle Exocytosis Are Required in the PG for Normal Developmental Progression

(A) Schematic illustration of components involved in calcium-regulated vesicle exocytosis in the PG cells. An unknown GPCR is shown in pink, SNARE complex proteins are shown as green ovals, and E is depicted as purple filled circles.

(B) Knockdown of the regulatory components for secretory vesicle exocytosis in the PG causes polyphasic growth arrest. Percentages of developmentally arrested larvae for each genotype are shown. Percentages of larvae arrested at first or second instar (L1/L2) are indicated in light gray, whereas those arrested at third instar (L3) are shown in dark gray. E feeding is indicated by + E. Numbers of animals tested are shown on top of each bar. Insets are representative images of wandering control larvae (control, *phm22 >* or *phm22>dicer2*) and RNAi larvae arrested as L3 (RNAi). Scale bars, 1 mm.

(legend continued on next page)

we focused our analysis on components that are more specifically involved in calcium-regulated exocytosis. UNC-13 is a highly conserved, plasma membrane-associated presynaptic protein with calcium-binding domains. It interacts with the SNARE protein syntaxin and primes synaptic vesicles for fusion and is essential for calcium-regulated synaptic vesicle exocytosis (Aravamudan et al., 1999; Richmond et al., 2001). There are three *Drosophila* genes that encode UNC-13 family proteins, and knockdown of only *CG34349* produced the larval arrest phenotype (Figure 2B; Table 1). Likewise, one of the secretory Rab GTPases, Rab3, and its interacting molecule RIM, both of which are involved in secretory vesicle trafficking and priming (Fukuda, 2008; Graf et al., 2012; Müller et al., 2012), are also essential for proper function of PG cells (Figure 2; Table 1). Synaptotagmins, which form another class of calcium sensor proteins critical for vesicle exocytosis (Chapman, 2008; de Wit et al., 2009), were also tested for their necessity in the PG, and only Syt1 was found to be required for PG function (Figure 2; Table 1). The knockdown phenotypes of these regulatory exocytosis components were strikingly similar to that of *IP3R*; they all showed polyphasic larval developmental arrest and pupariation delay accompanied by overgrowth, both of which were rescued by E feeding (Figure 2). Moreover, unlike *Sec10* RNAi, the knockdown of these components in the PG did not disrupt the plasma membrane localization of the mCD8-GFP reporter, suggesting that constitutive membrane traffic is not altered (Figure S2).

To further validate that the calcium signaling pathway is active in the PG cells at the time of metamorphosis initiation, we monitored calcium dynamics in these cells using the genetically encoded calcium indicator, GCaMP5 (Akerboom et al., 2012). As illustrated in Figures 2D and 2E, two types of spontaneous activities were observed in the PG cells of wandering larvae. One consisted of major concentration changes throughout the entire volume of a PG cell, which we refer to as macro spikes (Figures 2D and 2E; Movie S1). The number of active cells within a gland varied significantly, and the dynamics of the calcium concentration observed also varied in amplitude, duration, and frequency on a cell-by-cell basis (Figure 2E). A second activity, which we refer to as micro spikes (Figure 2E), appeared to occur in a limited area on the cell surface and exhibited faster kinetics. When the PG-specific knockdown of *Plc21C* was performed with two distinct RNAi constructs, a significant decrease in the number of animals exhibiting calcium dynamics of either class was observed (Figure 2F), while RNAi of a random control gene (*gbb*) had no effect. These results support the notion that GPCR-mediated calcium signaling is occurring in the PG cells prior to metamorphosis.

In order to genetically validate the coupling of GPCR-regulated calcium release and vesicle exocytosis, an activated form of Gαq (*Gαq[Q203L]*) was expressed in the PG (Figure S1C). Interestingly, constitutive activation of Gαq pathway in the PG led to early larval lethality, suggesting that timely activation of GPCR signaling pathway is critical for proper larval development. Importantly, this early larval lethality was rescued by co-expressing RNAi constructs of the anticipated downstream components (*IP3R*, *CG34349*, *Rab3*, and *Syt1*; Figure S1C) and more larvae developed into L3 or beyond, consistent with the knockdown phenotypes of those downstream genes (Figures 1 and 2). Taken together, these results suggest that the GPCR-regulated calcium release through IP3R is indeed coupled with vesicle exocytosis in the PG cells.

Syt1-Positive Vesicles Accumulate in the PG upon Calcium Signaling Knockdown

In order to visualize putative secretory vesicles whose exocytosis is regulated by calcium signaling we expressed in the PG eGFP-tagged Syt1 (Syt-GFP), a widely used secretory vesicle marker in both neuronal and non-neuronal cells (Sugita et al., 2001; Zhang et al., 2002). In wild-type larvae, Syt-GFP labeled both the plasma membrane and a small number of vesicles in the PG (Figures 3A and 3C). It is known that Syt-GFP often labels the plasma membrane, depending on its expression level (Kanno and Fukuda, 2008). Interestingly, knocking down *IP3R* in the PG resulted in prominent accumulation of the Syt-GFP vesicles in the cytoplasm, especially in the areas adjacent to the plasma membrane (Figures 3B, 3D, and 3E). *IP3R* knockdown in the PG did not alter the gross morphology of the PG (Figures 3B and S2), although the size of each cell might be slightly increased, which is potentially coupled with the accumulation of vesicles in the cytoplasm. This accumulation of Syt-GFP vesicles at the plasma membrane was typically observed in the PGs of *IP3R* RNAi larvae during the extended third instar stage (i.e., after 140 hr AEL), at the time when E accumulation was observed in the CNS-RG complexes (Figure 1F). This observation is consistent with our hypothesis that E is loaded into Syt1-positive secretory vesicles in the PG and is released into the hemolymph via exocytosis triggered by calcium signaling.

Atet Is an ABC Transporter Present in Syt1-Positive Vesicles in the PG

If E indeed requires vesicle-mediated machinery to be released from the PG, there should be transporters on the vesicle surface that load E into the vesicles. There is an ATP-binding cassette (ABC) transporter, E23, which has been proposed to function

---

(C) Knockdown of the regulatory components for secretory vesicle exocytosis in the PG causes developmental delay. Developmental time to pupation among non-arrested larvae is shown. Data are represented as mean ± SEM of three to seven independent experiments. Insets are images of control pupae (control, *phm22 >* or *phm22>dicer2*), RNAi pupae (RNAi), and RNAi pupae rescued by E feeding (RNAi + E). Scale bars, 1 mm.

(D) GCaMP5 calcium imaging of the PG cells from wandering third instar *phm22>GCaMP5* larvae. Cumulative maximum intensity projection of a 5 min time-lapse imaging is shown. Colored circles indicate the regions of interest (ROIs) where macro calcium spikes were observed. Scale bar, 25 μm.

(E) Plot of mean signal intensity in the cells indicated in (D). The color of each plot corresponds to that of the ROIs in (D). Inset is an example plot of a micro spike from a different sample.

(F) Quantification of the PGs that presented either macro (dark gray) or micro (light gray) calcium spikes in the animals of different genotypes (control, *phm22>GCaMP5*; RNAi, *phm22 > GCaMP5, RNAi*). Numbers of animals observed are shown on top of each genotype. **p < 0.01; ***p < 0.001 from Fisher's exact test compared to control.

See also Figure S2 and Movie S1.

Table 1. RNAi Screening of Genes Involved in Membrane Traffic, Intracellular Calcium Signaling, or Regulated Vesicle Exocytosis

	Classification	Gene Name	CG Number	VDRC Transformant ID	Phenotype
Membrane traffic	synaptobrevin	*Syb*	CG12210	39770*, 102922*	developmental delay/arrest
		n-Syb	CG17248	44011, 49201, 104531	no
	exocyst	*Sec3*	CG3885	35806, 108085	no
		Sec5	CG8843	28873*	developmental delay/arrest
		Sec6	CG5341	22077*, 105836*	developmental delay/arrest
		Sec8	CG2095	45032*, 105653*	developmental delay/arrest
		Sec10	CG6159	N/A*[a]	developmental delay/arrest
		Sec15	CG7034	35161*, 105126*	developmental delay/arrest
		exo70	CG7127	27867, 103717	no
		exo84	CG6095	30111*, 108650*	developmental delay/arrest
Intracellular Ca^{2+} release	Gαq	*CG30054*	CG30054	4643*[c], 4644*, 102887	developmental delay/arrest
		Gαq	CG17759	50729, 105300	no
		CG17760	CG17760	42255, 52308, 107613	no
	phospholipase C	*Plc21C*	CG4574	26557*, 108395*[c]	developmental delay/arrest
		norpA	CG3620	105676	no
		PLCγ	CG4200	7173, 108593	no
	ER Ca^{2+} channel	*IP3R*	CG1063	6484*[c], 106982*	developmental delay/arrest
		RyR	CG10844	109631	no
Ca^{2+}-regulated exocytosis	UNC-13	*CG34349*	CG34349	31571*[c], 31573*, 107855	developmental delay/arrest
		UNC-13	CG2999	33606, 33609, 101383	no
		UNC-13-4A	CG32381	41835, 109304	no
	secretory Rab	*Rab3*	CG7576	100787*[c]	developmental delay/arrest
		Rab26	CG34410	43730, 101330	no
		Rab27	CG14791	31887[b], 35774[b]	no
	RIM	*RIM*	CG33547	48072*[c], 39384	developmental delay/arrest
	synaptotagmin	*Syt1*	CG3139	100608*[c], 8874	developmental delay/arrest
		Syt4	CG10047	33317	no
		Syt7	CG2381	24988	no
		Syt12	CG10617	47504, 47506, 110655	no
		Syt14	CG9778	11037	no
		Styα	CG5559	9303, 100957	no
		Sytβ	CG42333	30013, 103345	no

UAS-RNAi lines from Vienna Drosophila RNAi Center (VDRC) were crossed to *phm22>dicer2* to induce tissue-specific knockdown in the PG. Multiple RNAi lines were tested for each gene whenever available to minimize the false-positive results from off-target effects. *Lines that exhibited the developmental delay/arrest phenotype.
See also Figure S1.
[a]Kind gift from K. Broadie.
[b]Lines from Transgenic RNAi Project at Harvard Medical School (obtained from Bloomington Drosophila Stock Center).
[c]Phenotypic rescue with E feeding was confirmed.

as a 20E exporter to modulate the effective intracellular concentration of 20E in peripheral tissues in *Drosophila* (Hock et al., 2000). E23 is a member of the ABCG subfamily of ABC transporters, several of which in mammals have been shown to help efflux cholesterol and other types of steroids such as estrogens and their metabolites (Imai et al., 2003; Janvilisri et al., 2003; Klucken et al., 2000; Suzuki et al., 2003; Wang et al., 2004; Yu et al., 2005). Based on these previous findings, an in situ hybridization screening of ABCG transporter genes in *Drosophila* genome (Figure S3) was conducted to identify putative E transporters highly expressed in the PG. This resulted in the identifica-

tion of *Atet* and *CG4822*, both of which are highly expressed in the PG of third instar larvae (Figures 4A and S3). Of these two genes, only knockdown of *Atet* in the PG showed developmental defects indistinguishable from those of knockdown of calcium-regulated vesicle exocytosis genes (Figures 4B–4D) and were also rescued by E feeding.

We next expressed a fluorescent protein-tagged Atet in the PG to visualize its subcellular localization. This resulted in the labeling of both the plasma membrane and Syt1-positive vesicles (Figure 4E), suggesting that Atet could indeed be involved in the import of E into these vesicles. The signal of fluorescent
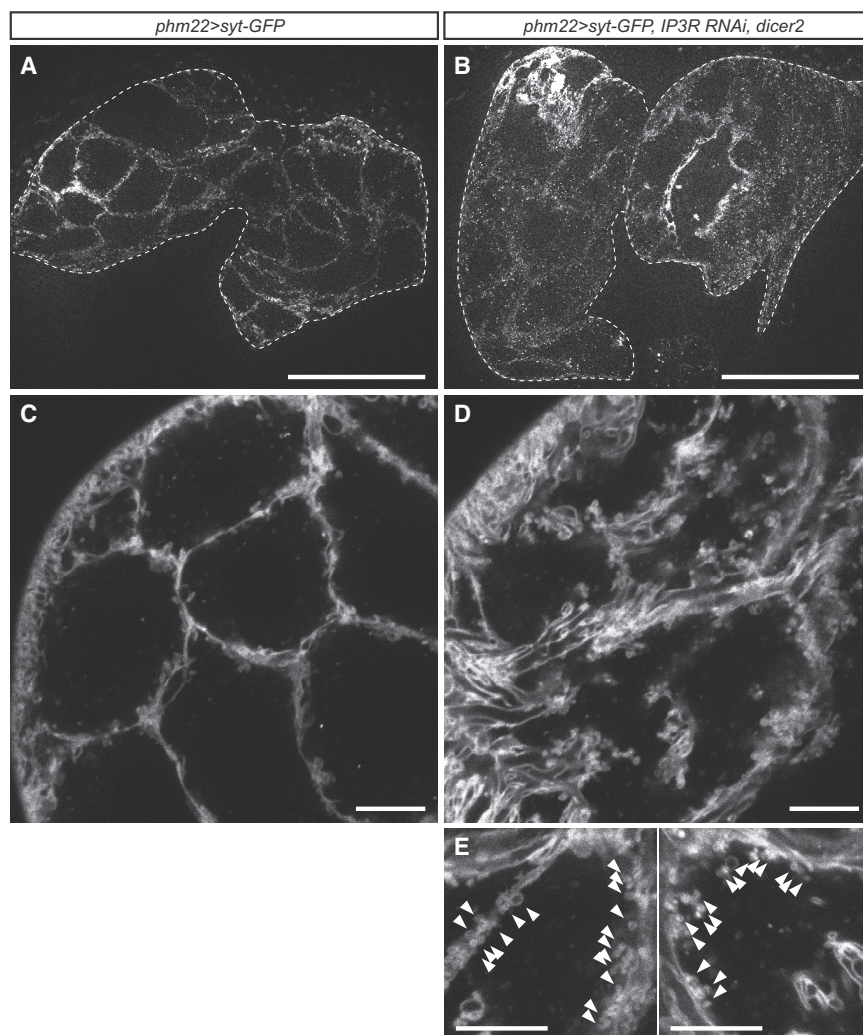
Figure 3. Syt-GFP Reveals the Presence of Vesicle-like Structures in the PG

(A) Representative image of the PG from a wandering control larva overexpressing Syt-GFP (*phm22>Syt-GFP*). The PG is surrounded by a dashed line. Scale bar, 100 μm.

(B) Representative image of the PG from a day 7 (∼150 hr AEL) *IP3R* RNAi larva overexpressing Syt-GFP (*phm22 > Syt-GFP, IP3R RNAi, dicer2*). The PG is surrounded by a dashed line. Scale bar, 100 μm.

(C) Confocal image of the PG from a wandering control larva overexpressing Syt-GFP. Scale bar, 10 μm.

(D) Confocal image of the PG from a day 7 (∼150 hr AEL) *IP3R* RNAi larva overexpressing Syt-GFP. Scale bar, 10 μm.

(E) Magnified view of the PG cells from day7 (∼150 hr AEL) *IP3R* RNAi larvae overexpressing Syt-GPF. Note aggregation of many small vesicle-like structures along the membrane (arrowheads). Scale bars, 10 μm.

protein-tagged CG4822, in contrast, did not co-localize with that of Atet in the PG (Figure S3E).

Atet Transports E across Membranes In Vitro in an ATP-Dependent Manner

In order to further examine if Atet is a critical transporter for loading E into PG secretory vesicles, we sought to develop an in vitro transport assay. We first analyzed the predicted membrane topology of Atet using Phobius, a transmembrane protein topology and signal peptide predictor program (Käll et al., 2007). To our surprise, Atet was predicted to have an extracellular N terminus, despite the fact that its ATP binding domain is on its N-terminal side (Figure 5A). The same result was obtained using two other independent algorithms (TMHMM version 2.0 [Krogh et al., 2001] and HMMTOP version 2.0 [Tusnády and Simon, 2001]), although the number of transmembrane domains differed between prediction algorithms. To determine the actual topology of Atet, we expressed an N-terminally HA-tagged Atet in Schneider 2 (S2) cells, a cell line derived from a primary culture of late stage *Drosophila* embryos, and immuno-stained the cells in both permeabilized

and non-permeabilized conditions (Figure 5B). Under permeable conditions, both the surface of the cells and the internal structures were stained, suggesting that a certain population of Atet proteins are localized on the plasma membrane as in the PG cells. Importantly, under non-permeable conditions, N-terminal staining of Atet was still detected on the surface of the cells, whereas the control E23 tagged at the intracellular C terminus was not detected without permeabilizing the cells (Figure 5B). These observations demonstrate that the N terminus of Atet is indeed located on the non-cytoplasmic side of the membranes.

Based on this atypical membrane topology of Atet, we designed an in vitro transport assay using S2 cell membrane vesicle preparations from cells transfected with Atet (Figure S4). A crude membrane preparation typically contains both inside-out and right-side-out vesicles. In a regular vesicular ABC transporter assay, only the activity of the transporters in the inside-out vesicle configuration are detected, since a typical ABC transporter in the right-side-out vesicles will have its ABC domain inside the vesicles and therefore unable to access the exogenously added ATP and transport substrate. Thus, activity is measured as the amount of substrate imported into the vesicles (Figure S4). In contrast, in the case of Atet, no net flux into vesicles would be expected upon addition of exogenous substrates since Atet is predicted to pump substrates in the opposite direction. Therefore, in our modified procedure, the substrate E was preloaded into vesicles during isolation and then ATP was added to assess the transporter activity as efflux rather than influx. (Figure S4). As shown in Figure 5C, addition of ATP significantly stimulated efflux of E from Atet containing vesicles while no transport was observed using E23, a putative E transporter with a 'normal'
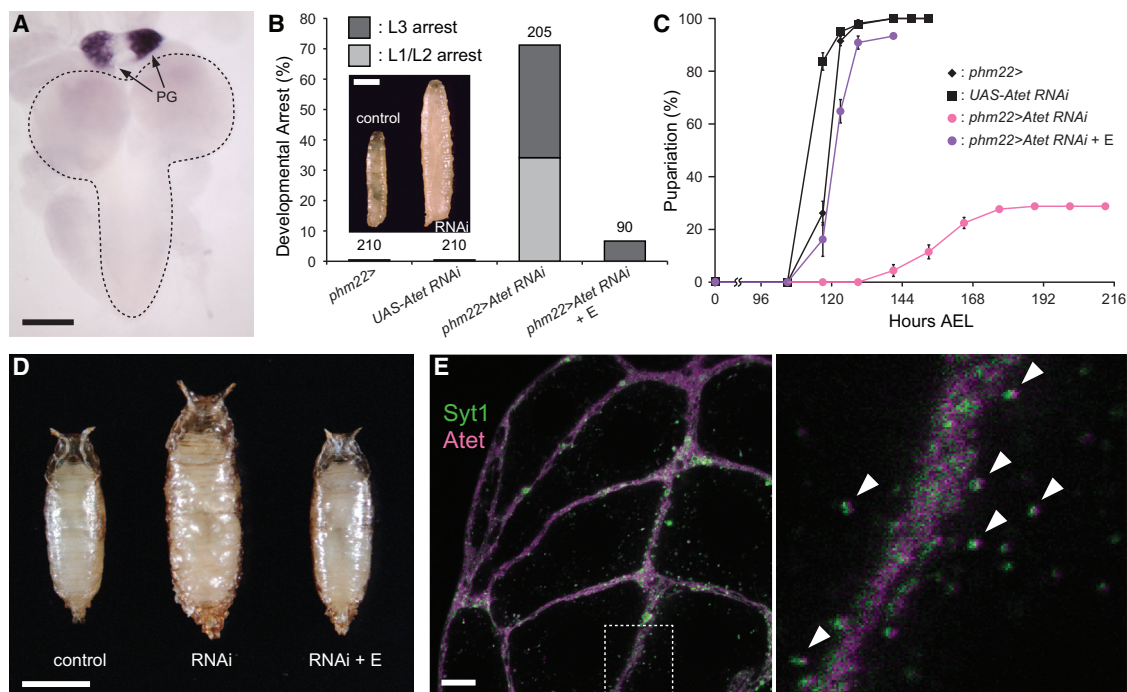
Figure 4. *Atet* Is Expressed in the PG and Required for Normal Developmental Progression

(A) *Atet* is highly expressed in the PG, as shown by in situ hybridization of the CNS-RG complex from a wandering larva with an *Atet* antisense probe. The PG components of the RG are indicated by arrows, and the CNS is surrounded by a dashed line. Scale bar, 100 μm.

(B) *Atet*-knockdown in the PG causes polyphasic growth arrest. Percentages of developmentally arrested larvae for each genotype are shown. Percentages of larvae arrested at first or second instar (L1/L2) are indicated in light gray, whereas those arrested at third instar (L3) are shown in dark gray. E feeding is indicated by + E. Numbers of animals tested are shown on top of each bar. Inset is a representative image of a wandering control larva (control, *phm22 >*) and an *Atet* RNAi larvae arrested as L3 (RNAi, *phm22 > Atet RNAi*). Scale bar, 1 mm.

(C) *Atet*-knockdown in the PG causes developmental delay. Developmental time to pupation among non-arrested larvae is shown. Data are represented as mean ± SEM of three to seven independent experiments.

(D) *Atet*-knockdown in the PG leads to the formation of overgrown pupae. Representative image of a control pupa (control; *phm22 >*), an *Atet*-knockdown pupa (RNAi, *phm22 > Atet RNAi*) and an *Atet*-knockdown pupa rescued by E feeding (RNAi + E, *phm22 > Atet RNAi* with E feeding). Scale bar, 1 mm.

(E) Syt1 and Atet co-localize in the PG. Representative confocal image of the PG from a wandering larva overexpressing Syt-GFP (green) and YPet-Atet (magenta). The square area surrounded by a dashed line in the left panel is magnified on the right. Vesicles labeled by both Syt-GFP and YPet-Atet are indicated by arrowheads. Scale bar, 10 μm.

See also Figure S3.

---

membrane configuration with respect to the ABC domain. These results demonstrate that Atet can indeed transport E from the cytoplasmic to non-cytoplasmic side of vesicle membranes, providing strong support for our vesicle-mediated E release model (Figure 6).

DISCUSSION

In the present study, we provide several lines of evidence demonstrating that the insect steroid hormone E is secreted from the PG not by simple diffusion, but rather through a calcium signaling-regulated vesicle fusion event. Below we discuss three major points of our findings: (1) Atet, an ABCG transporter, can facilitate E passage through membranes in an ATP-dependent manner, (2) GPCR-regulated calcium signaling in the PG promotes E release, and (3) the significance of steroid hormone release by vesicle exocytosis and its implication for other steroid hormone/cholesterol trafficking processes.

The ABCG Family Member Atet Is an E Transporter

Atet was originally cloned in *Drosophila* as an ABC transporter-encoding gene with unknown function. It was found to be highly expressed in embryonic trachea, leading to its name *ABC transporter expressed in trachea* or *Atet* (Kuwana et al., 1996). In our in situ hybridization experiment, however, we found little expression of *Atet* in embryonic trachea, but instead saw specific high level expression in the PG (Figure S3F), consistent with its expression pattern in the third instar larva (Figure 4A). Since we found that Atet has an atypical membrane topology (Figures 5A and 5B) and can transport E across membranes in vitro (Figure 5C), we propose renaming this gene *Atypical topology ecdysone transporter*, thereby retaining the *Atet* gene designation.

Atet belongs to the ABCG subfamily of ABC transporters, members of which in mammals have been shown to transport cholesterol as well as other steroids, such as estrogens and their metabolites, in many biological systems (Imai et al., 2003; Janvilisri et al., 2003; Klucken et al., 2000; Suzuki et al., 2003; Wang
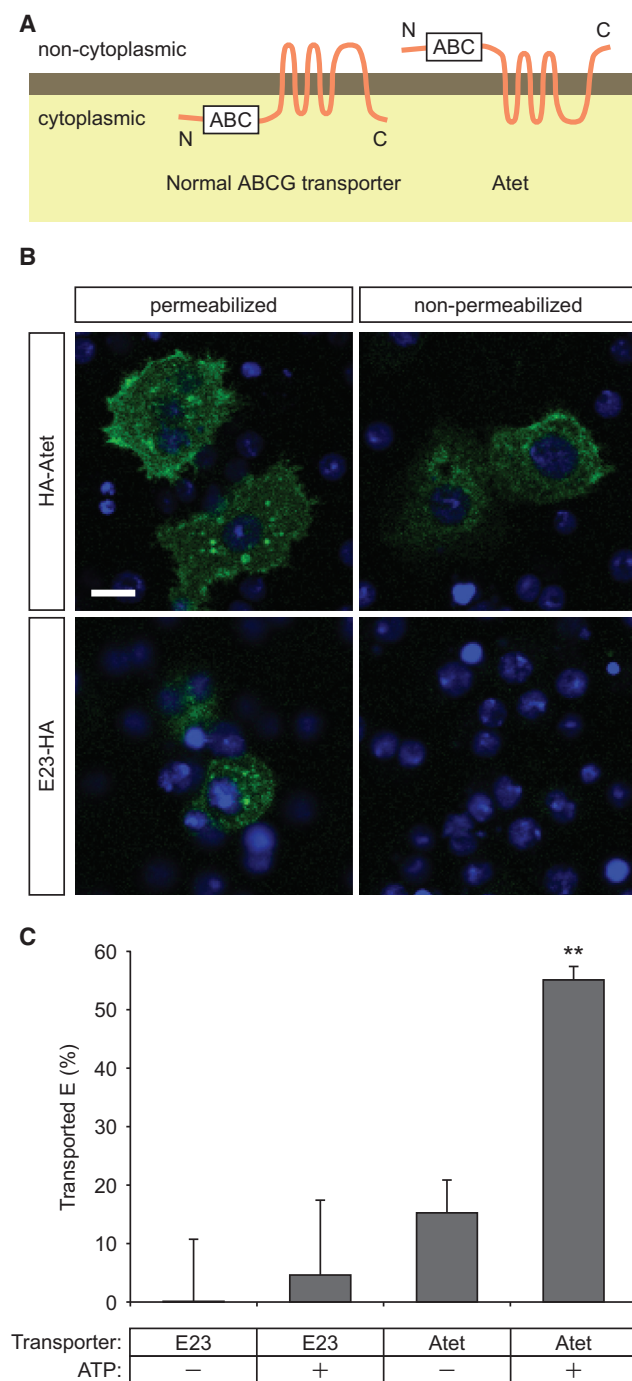
Figure 5. Atet Is an E Transporter

(A) Membrane topology of a normal ABC-type transporter (left) and of Atet (right) as predicted by membrane topology prediction servers. White boxes indicate the ATP bind cassette (ABC).

(B) Anti-HA staining of S2 cells overexpressing N-terminally tagged Atet (HA-Atet) or C-terminally tagged E23 (E23-HA) in permeabilized and non-permeabilized conditions. The C terminus of E23 was consistently predicted by prediction servers (Phobius, TMHMM and HMMTOP) to be intracellular and therefore used as a negative control.

(C) E transporting activity of Atet and E23 measured by modified vesicular transport assay. Data are represented as mean ± SEM of five to six inde-

et al., 2004; Yu et al., 2005). To our knowledge, however, the atypical membrane topology, with the N-terminal ABC domain on the non-cytoplasmic side of the membrane, has not been reported for any ABC transporter to date. However, this topology may have a strong advantage in facilitating tight control on E release by preventing Atet from functioning on the plasma membrane, due to the lack of ATP in extracellular space. This configuration therefore prevents E transport directly through the plasma membrane and confines it to a vesicle-mediated fusion process, although it requires a separate molecular mechanism to transport ATP into the secretory vesicles. This mechanism remains unclear at this point, but it may involve a specific transporter like the recently described VNUT/SLC17A9 (Sawada et al., 2008). In this context, it is interesting to note that the human Atet orthologs ABCG1 and ABCG4 (Figure S3A) are also strongly predicted by membrane topology algorithms to position their N-terminal ABC domain on the non-cytoplasmic side. These transporters mediate cellular cholesterol efflux (Wang et al., 2004) and have recently been shown to work not on the plasma membrane but in intracellular endosomes (Tarling and Edwards, 2011). Clearly, additional studies on the membrane topology of ABCG transporters are warranted.

Separate Signaling Pathways Likely Regulate E Production and Release

The results of our RNAi screening (Table 1) demonstrate that CG30054, a Gαq subunit, and Plc21C, a PLCβ class enzyme, are both required for proper PG function. These findings strongly implicate the existence of an unknown GPCR and cognate ligand as mediators of the calcium signaling event that we suggest stimulates E release from the PG. On the other hand, we know that the PTTH receptor is Torso, a receptor tyrosine kinase (Rewitz et al., 2009) and its primary role is to promote E production by inducing E biosynthetic enzyme gene transcription. These observations suggest that, at least in *Drosophila*, E production and release are likely regulated separately. This machinery might help the GPCR ligand to generate large pulses of steroid in a timely fashion. The identification of the GPCR as well as its ligand is necessary to further pursue this possibility.

Significance of Vesicle-Mediated E Release and Its Implication for Other Processes

The mechanism of steroid hormone transit through lipid membranes has not been well studied and in many physiology textbooks the issue is not even discussed. When this topic is mentioned, the explanation most often given is that they can freely diffuse through lipid membranes (Raven and Johnson, 2002; Sherwood, 2011; White and Porterfield, 2012). Despite this prevailing assumption, there are only a few reports where such transbilayer transfer of steroids by free diffusion has been analyzed. In one theoretical study, it was shown in silico that a free energy of solvation-based mechanism can produce rapid flux of estradiol, testosterone, and progesterone through

pendent experiments. **$p < 0.01$ compared to control (0% transport) from Student's t test.

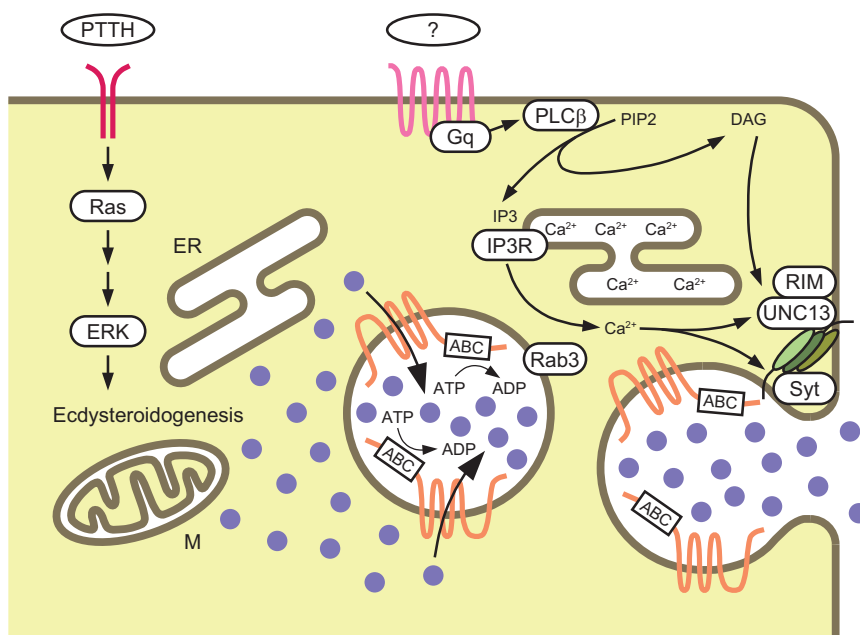See also Figure S4 for the details of the modified vesicular transport assay.

Figure 6. Schematic Illustration of the Vesicle-Mediated E Release Model
Torso is shown in red, an unknown GPCR is shown in pink, and Atet is shown in orange. E is depicted as purple filled circles. ER, endoplasmic reticulum; M, mitochondria.

one, testosterone, and estradiol are significantly more hydrophobic than E. Therefore, the free energy of solvation into a lipid bilayer of E is likely to be much more positive than for sex steroids; this may preclude the use of a simple diffusion mechanism for E. In this respect, E is more similar to bile acids, which are also highly hydrophilic and need active transporters to traverse lipid bilayers (Dawson et al., 2009). Thus, depending on their specific physiochemical properties, different steroids might use either simple passive diffusion through the plasma membrane, active transporters or some combination of these mechanisms.

a simple membrane in concordance with measured rates (Oren et al., 2004). However, it is well known that steroid hormone transport across membranes can indeed be an active process in some situations: there are a number of reports on transporter involvement in either uptake or elimination of steroid hormones in eukaryotes ranging from yeast to human (Hock et al., 2000; Janvilisri et al., 2003; Kralli et al., 1995; Kralli and Yamamoto, 1996; Mahé et al., 1996). These reports are suggestive enough to rationalize a potential mechanism that incorporates steroid hormones into secretory vesicles, which enables regulated secretion of steroid hormones from steroidogenic tissues.

Historically, the possibility of vesicle-mediated steroid hormone release has been examined using ultrastructural and biochemical approaches in multiple biological systems, including the corpus luteum in sheep (Gemmell and Stacy, 1979; Gemmell et al., 1974; Higuchi et al., 1976; Sawyer et al., 1979). The proposed vesicle-mediated progesterone release from the sheep corpus luteum, however, was later challenged, since the peptide oxytocin was shown to be present in dense granules by immuno-EM methods (Theodosis et al., 1986) and release of oxytocin and progesterone responded differently to various secretagogues (Hirst et al., 1986). Since that time, studies investigating the possibility of vesicle-mediated steroid release in any biological system have rarely been reported. One relevant and intriguing set of studies, however, involved ultrastructural localization of E in the PG of the waxworm *Galleria mellonella* (Birkenbeil, 1983; Birkenbeil et al., 1979) using immuno-EM methods. These studies suggested that E in the PG is concentrated into what appear to be secretory granules that fuse with the plasma membrane, but once again no follow up studies have been reported in the literature.

In considering the various models for steroid passage through membranes, it is important to note that steroids such as progester-

In summary, our work provides strong evidence that E is released from the PG by calcium-stimulated, vesicle-mediated exocytosis. Therefore, we suggest that the prevailing "free diffusion" model of steroid hormone secretion needs to be reconsidered. It also follows that if E uses an active export process, then the import of many hormones, in particular 20E, is also likely controlled by transporters. Given the diversity of physiological processes regulated by steroid hormones, additional characterization of the mechanisms responsible for their import and export from various cell types and tissues will have significant impact on both basic and clinical aspects of steroid hormone physiology.

EXPERIMENTAL PROCEDURES

Fly Stocks

All flies were raised at 25°C on standard medium under 12 hr/12 hr light/dark cycle. Aside from the control strains of *yw* and *w*, transgenic flies used in the figures are as follows: *phm22-Gal4* (Ou et al., 2011), *UAS-IP3R RNAi* (#6484, VDRC), *UAS-dicer2* (#60008 and #60009, VDRC), *UAS-Ras^V12* (#4847, BDSC), *UAS-CG34349 RNAi* (#31571, VDRC), *UAS-Rab3 RNAi* (100787, VDRC), *UAS-Syt1 RNAi* (#100608, VDRC), *UAS-GCaMP5* (#42038, BDSC), *UAS-gbb RNAi* (#5562R-1, NIG-FLY) *UAS-Gαq[Q203L]* (#30743, BDSC), *UAS-Syt-GFP* (#6926, BDSC), *UAS-Atet RNAi* (#42750, VDRC), *UAS-Sec10 RNAi* (a gift from K. Broadie) (Andrews et al., 2002), *OK72-Gal4* (#6486, BDSC), and *UAS-CG4822 RNAi* (#42730 and #105922, VDRC). All the other RNAi lines and sources used in this study are shown in Table 1. cDNA clones RE01860 and SD07027 from *Drosophila* Genomics Resource Center were used to tag *Atet* and *CG4822* with YPet or CyPet at N termini using the recombineering-mediated method (Venken et al., 2008). These products were then cloned into pUAST and transgenic flies were generated by BestGene.

Pupariation Timing and Developmental Arrest Phenotype Analyses

Synchronized newly hatched first instar larvae were placed on standard medium at 25°C, 25–30 larvae per vial, and pupariation timing of each individual was scored periodically. Developmentally arrested larvae were scored individually by checking their body and mouth hook characteristics. E feeding rescue was performed by adding 0.2 mg/ml (final concentration) of E into the medium. Larvae

were transferred into E-free medium during the early third instar stage (80–104 hr AEL) to avoid the potential detrimental effect of continuous E feeding.

Extraction of Ecdysteroids from CNS-RG Complexes and Hemolymph

Ten CNS-RG complexes were quickly dissected from larvae in PBS, briefly rinsed with PBS, and pooled in 300 μl of methanol on ice. The complexes were thoroughly homogenized by repeatedly passing through 23 gauge needles. After centrifugation at 4°C for 5 min, the supernatant was pooled on ice, while the pellet was re-extracted. The resulting extract (= 1 batch) was stored at −20°C until use. For hemolymph samples, 4 μl of hemolymph was collected from 10–20 larvae and mixed with 100 μl of methanol on ice. After vortexing, samples were centrifuged at 4°C for 5 min, and the resulting supernatant (= 1 batch) was stored at −20°C until use.

Separation of E and 20E Using HPLC

For each HPLC experiment, three batches of CNS-RG extract or one batch of hemolymph sample was partially evaporated with a SpeedVac concentrator and diluted with water to make the methanol concentration 30% or lower. After centrifugation at room temperature for 10 min to remove precipitates, the aqueous solution was applied onto a Vydac 218TP C18 column (4.6 × 250 mm; Grace) using Amersham Biosciences P-900 pump. The elution was performed with a linear gradient of 30%–50% methanol over 20 min at the flow rate of 1 ml/min. The fractions were collected every 30 s, and the amount of ecdysteroids in each fraction was determined by ecdysteroid ELISA using half the amount (250 μl) of each fraction. The residual half of each fraction corresponding to E (42.5%–44.5% methanol = 1 ml) and 20E (36%–38% methanol = 1 ml) were pooled and stored at −20°C until ecdysteroid quantification. This experiment was repeated three times for each biological sample.

Ecdysteroid ELISA

The sample solutions were dried with a SpeedVac concentrator and dissolved in EIA buffer (100 mM phosphate solution [pH 7.4], containing 0.1% BSA, 400 mM NaCl, 1 mM EDTA, and 0.01% NaN₃). 20E AChE tracer (#482200), 20E EIA antiserum (#482202), precoated (mouse anti-rabbit IgG) EIA 96-well plates (#400007), and Ellman's reagent (#400050) were all purchased from Cayman Chemical. The assay was performed according to the manufacturer's instructions using synthetic E or 20E (Sigma-Aldrich) as standards.

GCaMP5 Imaging and Data Analysis

Wandering third instar larvae were pinned down to a Sylgard dish and dissected in HL3 saline (2.5 mM Ca^{2+}) along the dorsal midline. All tissues with the exception of imaginal discs, CNS-RG complex and body wall musculature were removed. For the quantification of the PGs that presented calcium spikes (Figure 2F), the samples were treated with 1 μM tetrodotoxin in HL3 saline to block muscle contraction and stabilize the preparation. The samples were imaged in a Nikon FN1 microscope equipped with an A1R confocal scan head and a CFI75 Apo LWD 25× 1.1 NA water dipping objective. The PG was located using transmitted light and the zoom level was adjusted to include the whole PG within the field of view, which resulted in the pixel size of 0.25–0.33 μm. GCaMP5 was imaged with a 488 nm laser at constant power for all samples with the emission window of 500–550 nm. The experimental animals were coded so the imager didn't know the genotypes analyzed. The time-lapse runs were analyzed for the presence of macro and micro spikes in calcium levels (GCaMP5 intensity) within the PG cells before decoding the genotypes. All image processing and analysis was performed with the Nikon NIS-Elements software package.

Fluorescence Microscopy

To avoid disruption of vesicle structures by sample fixation, the PGs were dissected and mounted in Schneider's insect medium (Sigma-Aldrich) and imaged immediately with fluorescence microscopy. All confocal images were acquired using a Zeiss LSM 710 (Carl Zeiss). A lambda scan was performed and the signals of Syt-GFP and YPet-Atet were linearly unmixed using ZEN 2009 software. For observation of the entire PG in Figures 3A and 3B, Zeiss Axio Imager M2 equipped with ApoTome.2 and Plan-Apochromat 20× 0.8 NA objective (Carl Zeiss) was used.

In Situ Hybridization

In situ hybridization with DIG-labeled RNA antisense probe was performed as previously described (Chávez et al., 2000). cDNA clones used to generate antisense probes are shown in Figure S3A.

Immunocytochemistry

cDNA clones RE01860 and RE53253 from *Drosophila* Genomics Resource Center were used, respectively, to generate HA-tagged Atet and E23 constructs. A nonsense mutation found in the E23 clone RE53253 (bp 2766) was corrected by site-directed mutagenesis (QuikChange Kit; Agilent Technologies) using following primers: 5′-GGCCCAGCACCTGGTGTGGTGTGCC GCGGACTCGCAGTCC-3′; 5′-GGACTGCGAGTCCGCGGCACACCACACCA GGTGCTGGGCC-3′. An *Xba*I site was introduced downstream of the start codon (Atet) or upstream of the stop codon (E23) and was used to insert a triple HA epitope. These products were then cloned into pBRAcpA expression vector for transient transfection into S2 cells (Rewitz et al., 2009). After transfection, cells were grown in serum-free M3 medium (Sigma-Aldrich) for 4 days and attached to concanavalin A-coated slides overnight at 25°C. Membrane permeabilization was performed with 0.1% Triton X-100 in PBS for 15 min at room temperature. Cells were stained with rat monoclonal anti-HA antibody 3F10 (1:500; Roche Applied Science) followed by Alexa Fluor 488 Goat Anti-Rat IgG antibody (1:1,000; Molecular Probes) pre-absorbed with S2 cells overnight at 4°C. Cells were briefly treated with PBS containing DAPI before mounted in Vectashield (Vector Laboratories). Images were acquired using a Zeiss LSM 710 equipped with C-Apochromat 40× 1.2 NA water immersion and alpha Plan-Apochromat 100× 1.46 NA oil immersion objectives (Carl Zeiss).

Modified Vesicular Transport Assay

S2 cells were grown in 10-cm Petri dishes in serum-free M3 medium (Sigma-Aldrich) for 4 days after transfection with pBRAcpA empty vector (control), pBRAcpA-Atet, or pBRAcpA-E23-HA. After pelleting by brief centrifugation at room temperature, the cells from two dishes were suspended in 4 ml ice-cold extraction buffer (50 mM HEPES-KOH [pH 7.5], containing 400 mM sucrose, supplemented with cOmplete, Mini, EDTA-free protease inhibitor cocktail, Roche Applied Science) and sonicated for 30 s with a probe sonicator two times on ice. The cell lysates were then centrifuged at 1,000 × g for 5 min at 4°C to pellet cell debris, and the resultant supernatant was ultracentrifuged at 108,000 × g for 30 min at 4°C. The pellet (membrane fraction) was suspended in 1.6 ml Assay buffer (50 mM HEPES-KOH [pH 7.5], containing 10 mM MgCl₂) and sonicated for 30 s with a probe sonicator on ice. The suspension was then immediately aliquoted (400 μl each) into four 1.5-ml tubes containing 50 μl E solution (100 μg/ml in Assay buffer) to incorporate E into vesicles. After 15 min, 50 μl of ATP solution (Sigma-Aldrich, 10 mM in Assay buffer) was added to each tube to give a final concentration of 1mM, and the transport assay was initiated. After 1 hr incubation at 25°C, the vesicles were collected by filtrating through Whatman GF/A Glass microfiber filters (GE Healthcare). After washing three times with 1 ml 50 mM HEPES-KOH [pH 7.5], the filters were vigorously vortexed in 800 μl of methanol to extract E, and the E amount was determined by ecdysteroid ELISA.

REFERENCES

Akerboom, J., Chen, T.W., Wardill, T.J., Tian, L., Marvin, J.S., Mutlu, S., Calderón, N.C., Esposti, F., Borghuis, B.G., Sun, X.R., et al. (2012). Optimization of a GCaMP calcium indicator for neural activity imaging. J. Neurosci. *32*, 13819–13840.

Andrews, H.K., Zhang, Y.Q., Trotta, N., and Broadie, K. (2002). Drosophila sec10 is required for hormone secretion but not general exocytosis or neurotransmission. Traffic *3*, 906–921.

Aravamudan, B., Fergestad, T., Davis, W.S., Rodesch, C.K., and Broadie, K. (1999). Drosophila UNC-13 is essential for synaptic transmission. Nat. Neurosci. *2*, 965–971.

Birkenbeil, H. (1983). Ultrastructural and immunocytochemical investigation of ecdysteroid secretion by the prothoracic gland of the waxmoth Galleria mellonella. Cell Tissue Res. *229*, 433–441.

Birkenbeil, H., Eckert, M., and Gersch, M. (1979). Electron microscopical-immunocytochemical evidence of ecdysteroids in the prothoracic gland of Galleria mellonella. Cell Tissue Res. *200*, 285–290.

Cáceres, L., Necakov, A.S., Schwartz, C., Kimber, S., Roberts, I.J., and Krause, H.M. (2011). Nitric oxide coordinates metabolism, growth, and development via the nuclear receptor E75. Genes Dev. *25*, 1476–1485.

Chapman, E.R. (2008). How does synaptotagmin trigger neurotransmitter release? Annu. Rev. Biochem. *77*, 615–641.

Chávez, V.M., Marqués, G., Delbecque, J.P., Kobayashi, K., Hollingsworth, M., Burr, J., Natzle, J.E., and O'Connor, M.B. (2000). The Drosophila disembodied gene controls late embryonic morphogenesis and codes for a cytochrome P450 enzyme that regulates embryonic ecdysone levels. Development *127*, 4115–4126.

Dawson, P.A., Lan, T., and Rao, A. (2009). Bile acid transporters. J. Lipid Res. *50*, 2340–2357.

de Wit, H., Walter, A.M., Milosevic, I., Gulyás-Kovács, A., Riedel, D., Sørensen, J.B., and Verhage, M. (2009). Synaptotagmin-1 docks secretory vesicles to syntaxin-1/SNAP-25 acceptor complexes. Cell *138*, 935–946.

Fukuda, M. (2008). Regulation of secretory vesicle traffic by Rab small GTPases. Cell. Mol. Life Sci. *65*, 2801–2813.

Gemmell, R.T., and Stacy, B.D. (1979). Granule secretion by the luteal cell of the sheep: the fate of the granule membrane. Cell Tissue Res. *197*, 413–419.

Gemmell, R.T., Stacy, B.D., and Thorburn, G.D. (1974). Ultrastructural study of secretory granules in the corpus luteum of the sheep during the estrous cycle. Biol. Reprod. *11*, 447–462.

Ghayee, H.K., and Auchus, R.J. (2007). Basic concepts and recent developments in human steroid hormone biosynthesis. Rev. Endocr. Metab. Disord. *8*, 289–300.

Graf, E.R., Valakh, V., Wright, C.M., Wu, C., Liu, Z., Zhang, Y.Q., and DiAntonio, A. (2012). RIM promotes calcium channel accumulation at active zones of the Drosophila neuromuscular junction. J. Neurosci. *32*, 16586–16596.

Higuchi, T., Kaneko, A., Abel, J.H., Jr., and Niswender, G.D. (1976). Relationship between membrane potential and progesterone release in ovine corpora lutea. Endocrinology *99*, 1023–1032.

Hirst, J.J., Rice, G.E., Jenkin, G., and Thorburn, G.D. (1986). Secretion of oxytocin and progesterone by ovine corpora lutea in vitro. Biol. Reprod. *35*, 1106–1114.

Hock, T., Cottrill, T., Keegan, J., and Garza, D. (2000). The E23 early gene of Drosophila encodes an ecdysone-inducible ATP-binding cassette transporter capable of repressing ecdysone-mediated gene activation. Proc. Natl. Acad. Sci. USA *97*, 9519–9524.

Huang, X., Warren, J.T., and Gilbert, L.I. (2008). New players in the regulation of ecdysone biosynthesis. J. Genet. Genomics *35*, 1–10.

Imai, Y., Asada, S., Tsukahara, S., Ishikawa, E., Tsuruo, T., and Sugimoto, Y. (2003). Breast cancer resistance protein exports sulfated estrogens but not free estrogens. Mol. Pharmacol. *64*, 610–618.

Janvilisri, T., Venter, H., Shahi, S., Reuter, G., Balakrishnan, L., and van Veen, H.W. (2003). Sterol transport by the human breast cancer resistance protein (ABCG2) expressed in Lactococcus lactis. J. Biol. Chem. *278*, 20645–20651.

Käll, L., Krogh, A., and Sonnhammer, E.L. (2007). Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server. Nucleic Acids Res. *35*, W429–W432.

Kanno, E., and Fukuda, M. (2008). Increased plasma membrane localization of O-glycosylation-deficient mutant of synaptotagmin I in PC12 cells. J. Neurosci. Res. *86*, 1036–1043.

Klucken, J., Büchler, C., Orsó, E., Kaminski, W.E., Porsch-Ozcürümez, M., Liebisch, G., Kapinsky, M., Diederich, W., Drobnik, W., Dean, M., et al. (2000). ABCG1 (ABC8), the human homolog of the Drosophila white gene, is a regulator of macrophage cholesterol and phospholipid transport. Proc. Natl. Acad. Sci. USA *97*, 817–822.

Kralli, A., and Yamamoto, K.R. (1996). An FK506-sensitive transporter selectively decreases intracellular levels and potency of steroid hormones. J. Biol. Chem. *271*, 17152–17156.

Kralli, A., Bohen, S.P., and Yamamoto, K.R. (1995). LEM1, an ATP-binding-cassette transporter, selectively modulates the biological potency of steroid hormones. Proc. Natl. Acad. Sci. USA *92*, 4701–4705.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.

Kuwana, H., Shimizu-Nishikawa, K., Iwahana, H., and Yamamoto, D. (1996). Molecular cloning and characterization of the ABC transporter expressed in Trachea (ATET) gene from Drosophila melanogaster. Biochim. Biophys. Acta *1309*, 47–52.

Langevin, J., Morgan, M.J., Sibarita, J.B., Aresta, S., Murthy, M., Schwarz, T., Camonis, J., and Bellaïche, Y. (2005). Drosophila exocyst components Sec5, Sec6, and Sec15 regulate DE-Cadherin trafficking from recycling endosomes to the plasma membrane. Dev. Cell *9*, 365–376.

Li, L., and Chin, L.S. (2003). The molecular machinery of synaptic vesicle exocytosis. Cell. Mol. Life Sci. *60*, 942–960.

Mahé, Y., Lemoine, Y., and Kuchler, K. (1996). The ATP binding cassette transporters Pdr5 and Snq2 of Saccharomyces cerevisiae can mediate transport of steroids in vivo. J. Biol. Chem. *271*, 25167–25172.

Miller, W.L. (2013). Steroid hormone synthesis in mitochondria. Mol. Cell. Endocrinol. *379*, 62–73.

Müller, M., Liu, K.S., Sigrist, S.J., and Davis, G.W. (2012). RIM controls homeostatic plasticity through modulation of the readily-releasable vesicle pool. J. Neurosci. *32*, 16574–16585.

Murthy, M., Garza, D., Scheller, R.H., and Schwarz, T.L. (2003). Mutations in the exocyst component Sec5 disrupt neuronal membrane traffic, but neurotransmitter release persists. Neuron *37*, 433–447.

Oren, I., Fleishman, S.J., Kessel, A., and Ben-Tal, N. (2004). Free diffusion of steroid hormones across biomembranes: a simplex search with implicit solvent model calculations. Biophys. J. *87*, 768–779.

Ou, Q., Magico, A., and King-Jones, K. (2011). Nuclear receptor DHR4 controls the timing of steroid hormone pulses during Drosophila development. PLoS Biol. *9*, e1001160.

Raven, P.H., and Johnson, G.B. (2002). The endocrine system. In Biology (McGraw-Hill), pp. 1125–1146.

Rewitz, K.F., Yamanaka, N., Gilbert, L.I., and O'Connor, M.B. (2009). The insect neuropeptide PTTH activates receptor tyrosine kinase torso to initiate metamorphosis. Science *326*, 1403–1405.

Richmond, J.E., and Broadie, K.S. (2002). The synaptic vesicle cycle: exocytosis and endocytosis in Drosophila and C. elegans. Curr. Opin. Neurobiol. *12*, 499–507.

Richmond, J.E., Weimer, R.M., and Jorgensen, E.M. (2001). An open form of syntaxin bypasses the requirement for UNC-13 in vesicle priming. Nature *412*, 338–341.

Rizo, J., and Rosenmund, C. (2008). Synaptic vesicle fusion. Nat. Struct. Mol. Biol. *15*, 665–674.

Sapolsky, R.M., Romero, L.M., and Munck, A.U. (2000). How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. Endocr. Rev. *21*, 55–89.

Sawada, K., Echigo, N., Juge, N., Miyaji, T., Otsuka, M., Omote, H., Yamamoto, A., and Moriyama, Y. (2008). Identification of a vesicular nucleotide transporter. Proc. Natl. Acad. Sci. USA *105*, 5683–5686.

Sawyer, H.R., Abel, J.H., Jr., McClellan, M.C., Schmitz, M., and Niswender, G.D. (1979). Secretory granules and progesterone secretion by ovine corpora lutea in vitro. Endocrinology *104*, 476–486.

Sherwood, L. (2011). Principles of neural and hormonal communication. In Fundamentals of Human Physiology (Cengage Learning), pp. 71–105.

Shortridge, R.D., and McKay, R.R. (1995). Invertebrate phosphatidylinositol-specific phospholipases C and their role in cell signaling. Invert. Neurosci. *1*, 199–206.

Sisk, C.L., and Foster, D.L. (2004). The neural basis of puberty and adolescence. Nat. Neurosci. *7*, 1040–1047.

Sorrentino, V., Barone, V., and Rossi, D. (2000). Intracellular Ca(2+) release channels in evolution. Curr. Opin. Genet. Dev. *10*, 662–667.

Sudhof, T.C. (2004). The synaptic vesicle cycle. Annu. Rev. Neurosci. *27*, 509–547.

Sugita, S., Han, W., Butz, S., Liu, X., Fernández-Chacón, R., Lao, Y., and Südhof, T.C. (2001). Synaptotagmin VII as a plasma membrane Ca(2+) sensor in exocytosis. Neuron *30*, 459–473.

Suzuki, M., Suzuki, H., Sugimoto, Y., and Sugiyama, Y. (2003). ABCG2 transports sulfated conjugates of steroids and xenobiotics. J. Biol. Chem. *278*, 22644–22649.

Talamillo, A., Sánchez, J., Cantera, R., Pérez, C., Martín, D., Caminero, E., and Barrio, R. (2008). Smt3 is required for Drosophila melanogaster metamorphosis. Development *135*, 1659–1668.

Tarling, E.J., and Edwards, P.A. (2011). ATP binding cassette transporter G1 (ABCG1) is an intracellular sterol transporter. Proc. Natl. Acad. Sci. USA *108*, 19719–19724.

Theodosis, D.T., Wooding, F.B., Sheldrick, E.L., and Flint, A.P. (1986). Ultrastructural localisation of oxytocin and neurophysin in the ovine corpus luteum. Cell Tissue Res. *243*, 129–135.

Tusnády, G.E., and Simon, I. (2001). The HMMTOP transmembrane topology prediction server. Bioinformatics *17*, 849–850.

Venkatesh, K., and Hasan, G. (1997). Disruption of the IP3 receptor gene of Drosophila affects larval metamorphosis and ecdysone release. Curr. Biol. *7*, 500–509.

Venken, K.J., Kasprowicz, J., Kuenen, S., Yan, J., Hassan, B.A., and Verstreken, P. (2008). Recombineering-mediated tagging of Drosophila genomic constructs for in vivo localization and acute protein inactivation. Nucleic Acids Res. *36*, e114.

Wang, N., Lan, D., Chen, W., Matsuura, F., and Tall, A.R. (2004). ATP-binding cassette transporters G1 and G4 mediate cellular cholesterol efflux to high-density lipoproteins. Proc. Natl. Acad. Sci. USA *101*, 9774–9779.

Watanabe, S., Tani, T., Watanabe, S., and Seno, M. (1991). Transport of steroid hormones facilitated by serum proteins. Biochim. Biophys. Acta *1073*, 275–284.

White, B.A., and Porterfield, S.P. (2012). Introduction to the endocrine system. In Endocrine and Reproductive Physiology (Mosby/Elsevier), pp. 1–25.

Yamanaka, N., Rewitz, K.F., and O'Connor, M.B. (2013a). Ecdysone control of developmental transitions: lessons from Drosophila research. Annu. Rev. Entomol. *58*, 497–516.

Yamanaka, N., Romero, N.M., Martin, F.A., Rewitz, K.F., Sun, M., O'Connor, M.B., and Léopold, P. (2013b). Neuroendocrine control of Drosophila larval light preference. Science *341*, 1113–1116.

Yu, L., Gupta, S., Xu, F., Liverman, A.D., Moschetta, A., Mangelsdorf, D.J., Repa, J.J., Hobbs, H.H., and Cohen, J.C. (2005). Expression of ABCG5 and ABCG8 is required for regulation of biliary cholesterol secretion. J. Biol. Chem. *280*, 8742–8747.

Zhang, Y.Q., Rodesch, C.K., and Broadie, K. (2002). Living synaptic vesicle marker: synaptotagmin-GFP. Genesis *34*, 142–145.

Cell

Histone H3 Threonine Phosphorylation Regulates Asymmetric Histone Inheritance in the *Drosophila* Male Germline

Graphical Abstract
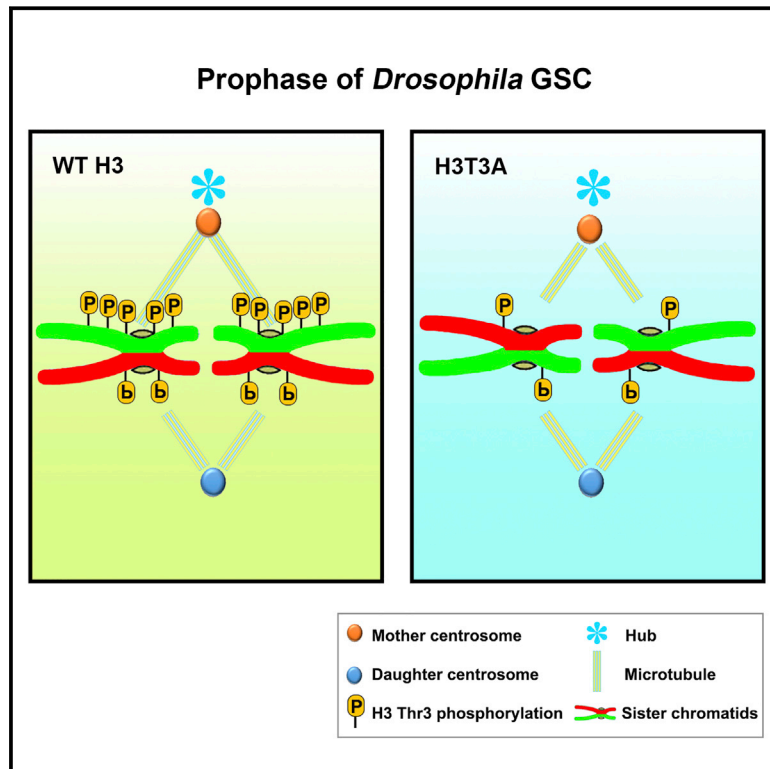
Authors

Jing Xie, Matthew Wooten, Vuong Tran, ..., Christine Simbolon, Eric Betzig, Xin Chen

Correspondence

xchen32@jhu.edu

In Brief

A transient mitosis-specific phosphate modification on histone H3 distinguishes pre-existing and newly synthesized histones and is required for the asymmetric segregation of sister chromatids—one enriched with new histones and the other with old—during stem cell division.

Highlights

- Pre-existing versus newly synthesized H3 are separable in prophase germline stem cell

- H3 threonine 3 phosphorylation distinguishes pre-existing from newly synthesized H3

- Both H3T3A and H3T3D mutations randomize H3 inheritance patterns

- Reducing H3T3 kinase Haspin enhances the H3T3A but suppresses the H3T3D phenotypes

CellPress

Article

Cell

Histone H3 Threonine Phosphorylation Regulates Asymmetric Histone Inheritance in the *Drosophila* Male Germline

Jing Xie,[1] Matthew Wooten,[1] Vuong Tran,[1,3] Bi-Chang Chen,[2,4] Caitlin Pozmanter,[1] Christine Simbolon,[1] Eric Betzig,[2] and Xin Chen[1,*]

[1]Department of Biology, Johns Hopkins University, Baltimore, MD 21218, USA
[2]HHMI, Janelia Research Campus, 19700 Helix Drive, Ashburn, VA 20147, USA
[3]Present address: Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North Seattle, Seattle, WA 98109, USA
[4]Present address: Research Center for Applied Science, Academia Sinica, Taipei, 11529, Taiwan
*Correspondence: xchen32@jhu.edu
http://dx.doi.org/10.1016/j.cell.2015.10.002

SUMMARY

A long-standing question concerns how stem cells maintain their identity through multiple divisions. Previously, we reported that pre-existing and newly synthesized histone H3 are asymmetrically distributed during *Drosophila* male germline stem cell (GSC) asymmetric division. Here, we show that phosphorylation at threonine 3 of H3 (H3T3P) distinguishes pre-existing versus newly synthesized H3. Converting T3 to the unphosphorylatable residue alanine (H3T3A) or to the phosphomimetic aspartate (H3T3D) disrupts asymmetric H3 inheritance. Expression of H3T3A or H3T3D specifically in early-stage germline also leads to cellular defects, including GSC loss and germline tumors. Finally, compromising the activity of the H3T3 kinase Haspin enhances the H3T3A but suppresses the H3T3D phenotypes. These studies demonstrate that H3T3P distinguishes sister chromatids enriched with distinct pools of H3 in order to coordinate asymmetric segregation of "old" H3 into GSCs and that tight regulation of H3T3 phosphorylation is required for male germline activity.

INTRODUCTION

Epigenetic phenomena are heritable changes in gene expression or function that can persist throughout many cell divisions without alterations in primary DNA sequences. By regulating differential gene expression, epigenetic processes are able to direct cells with identical genomes to become distinct cell types in humans and other multicellular organisms. However, with the exception of DNA methylation, little is known about the molecular pathways leading to epigenetic inheritance (Bonasio et al., 2010; Martin and Zhang, 2007).

Prior research has shown that epigenetic events play particularly important roles in ensuring both proper maintenance and differentiation of several stem cell populations. Many types of

adult stem cells undergo asymmetric cell division to generate a self-renewed stem cell and a daughter cell that will subsequently differentiate (Betschinger and Knoblich, 2004; Clevers, 2005; Inaba and Yamashita, 2012; Morrison and Kimble, 2006). Misregulation of this balance leads to many human diseases, ranging from cancer to tissue dystrophy to infertility. However, the mechanisms of stem cell epigenetic memory maintenance as well as how loss of this memory contributes to disease remain unknown.

Recently, we found that during the asymmetric division of the *Drosophila* male germline stem cell (GSC), the pre-existing histone 3 (H3) is selectively segregated to the self-renewed GSC daughter cell whereas newly synthesized H3 is enriched in the differentiating daughter cell known as a gonialblast (GB) (Tran et al., 2012) (Figure 1A). In contrast, the histone variant H3.3, which is incorporated in a replication-independent manner, does not exhibit such an asymmetric pattern. Furthermore, we found that asymmetric H3 inheritance occurs specifically in asymmetrically dividing GSCs, but not in the symmetrically dividing progenitor cells. These findings demonstrate that global asymmetric H3 histone inheritance possesses both molecular and cellular specificity. We proposed the following model to explain our findings.

First, the cellular specificity exhibited by the H3 histone suggests that global asymmetric histone inheritance occurs uniquely in a cell-type (GSC) where the mother cell must divide to produce two daughter cells each with a unique cell fate. Because this asymmetry is not observed in symmetrically dividing GB cells, we propose asymmetric histone inheritance to be a phenomenon specifically employed by GSCs to establish unique epigenetic identities in each of the two daughter cells. Second, as stated previously, a major difference between H3 and H3.3 is that H3 is incorporated to chromatin during DNA replication, while H3.3 variant is incorporated in a replication-independent manner. Because this asymmetric inheritance mode is specific to H3, we propose a two-step model to explain asymmetric H3 inheritance: (1) prior to mitosis, pre-existing and newly synthesized H3 are differentially distributed on the two sets of sister chromatids, and (2) during mitosis, the set of sister chromatids containing pre-existing H3 is segregated to GSCs, while the set of sister chromatids enriched with newly synthesized H3
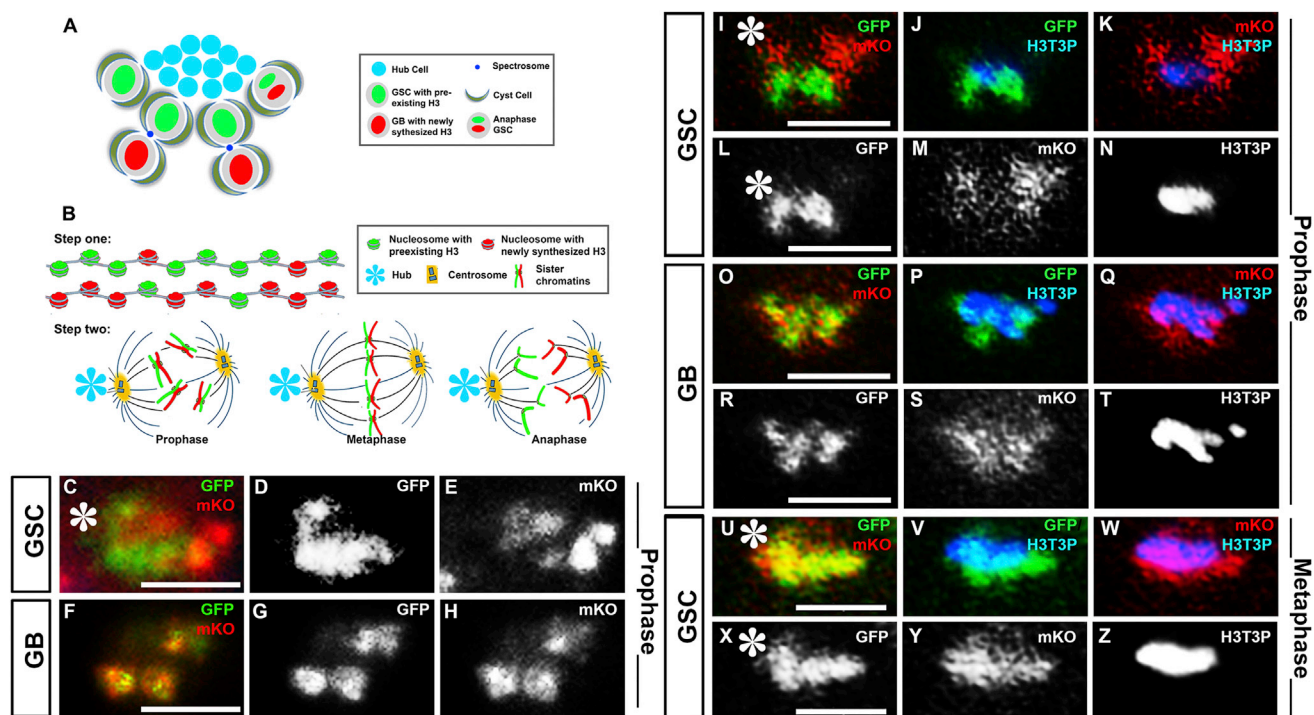
Cell

Figure 1. H3T3P Distinguishes Pre-existing H3-GFP from Newly Synthesized H3-mKO in Mitotic Male GSCs

(A) A visual representation of the *Drosophila* testis tip showing the asymmetric H3 inheritance during male GSC asymmetric cell division.

(B) A schematic diagram of a two-step model to explain how the asymmetric epigenome is established during S-phase (step one) and recognized followed by asymmetric segregation in M-phase (step two) GSC, adapted from Tran et al. (2013).

(C–E) A prophase GSC where GFP and mKO signals are separable.

(F–H) A prophase GB where GFP and mKO signals are overlapping.

(I–N) A prophase GSC where GFP and mKO signals are separable at some chromosomal region (I, L, and M). Immunostaining using anti-H3T3P (N) showed H3T3P co-localization more with GFP (J, L, and N) than with mKO (K, M, and N).

(O–T) A prophase GB where GFP and mKO signals are overlapping (O, R, and S) and no preference of H3T3P (T) with either GFP (P and R) or mKO (Q and S).

(U–Z) A metaphase GSC where GFP and mKO signals are indistinguishable (U, X, and Y), H3T3P (Z) overlaps with both GFP (V and X) and mKO (W and Y).

Asterisks in (C), (I), (L), (U), and (X), hub. Scale bars, 5 μm.

See also Figure S1.

is segregated to the GB that differentiates (Tran et al., 2012, 2013) (Figure 1B).

RESULTS

H3T3P Distinguishes Pre-existing H3 and Newly Synthesized H3 in Mitotic Male GSCs

To test our proposed two-step model, we used a temporally controlled dual-color system to precisely label pre-existing H3 with GFP and newly synthesized H3 with monomeric Kusabira-Orange (mKO) (Tran et al., 2012). Asymmetric segregation of H3-GFP and H3-mKO was clearly visualized in anaphase and telophase GSCs imaged during the second mitosis following heat-shock-induced switch from *H3-GFP*- to *H3-mKO*-coding sequence (Tran et al., 2012). Here, we show that H3-GFP and H3-mKO signals are already separable at some chromosomal region in prophase GSCs (Figures 1C–1E), likely in regions with less tight cohesion between sister chromatids. Such a separation was not detected in a control prophase GB (Figures 1F–1H). These results are consistent with the hypothesis that the differential distribution between

pre-existing H3-GFP and newly synthesized H3-mKO is established prior to mitosis in GSCs (Figure 1B, step one). By contrast, such a separation was not detected using a H3.3 dual-color transgene under the same heat-shock regime (Figure S1A), consistent with our previous report that H3.3 is inherited symmetrically (Tran et al., 2012).

When immunostaining experiments were performed using an antibody recognizing a mitosis-enriched phosphorylation at threonine 3 of H3 (H3T3P) (Dai et al., 2005; Polioudaki et al., 2004), the H3T3P signal (Figures 1J, 1K, and 1N) showed more co-localization with H3-GFP (Figures 1I and 1L) than with H3-mKO (Figures 1I and 1M) in prophase GSCs where separation between H3-GFP and H3-mKO could be visualized (Figures 1I, 1L, and 1M). By contrast, H3-GFP signals and H3-mKO signals were not separable in prophase GBs, (Figures 1O, 1R, and 1S) and H3T3P did not distinguish between them (Figures 1P, 1Q, and 1T). Furthermore, when sister chromatids congressed to the equator in metaphase GSCs, such a distinction became undetectable (Figures 1U–1Z), suggesting that H3T3P distinguishes sister chromatids enriched with pre-existing H3 from those enriched with newly synthesized H3 in prophase GSCs.
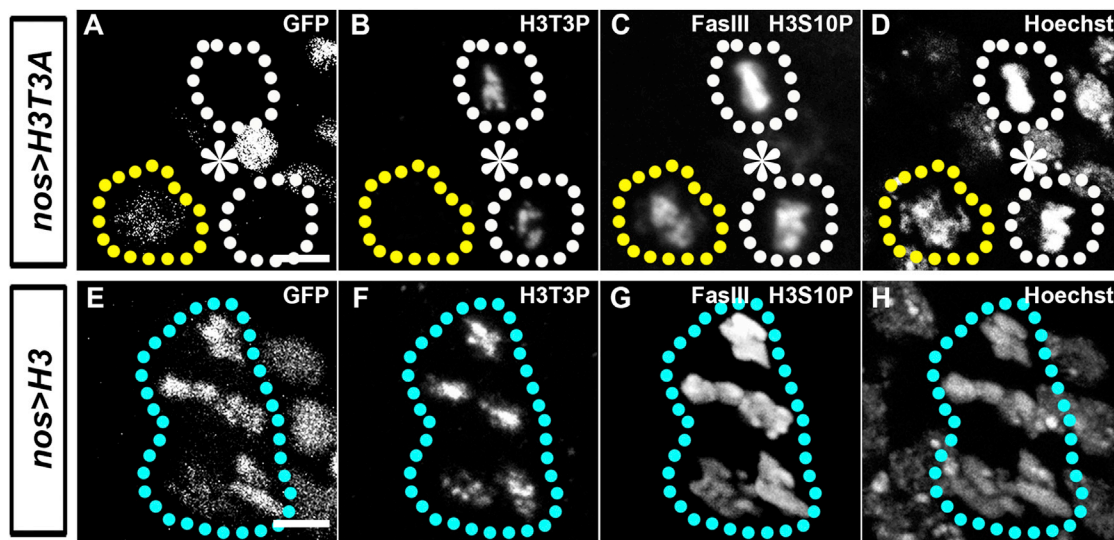
Cell

Figure 2. Expression of an H3T3A Transgene Greatly Reduces H3T3P in Mitotic Germ Cells

(A–D) Tip of a testis expressing *nos>H3T3A-GFP* stained with antibodies against a hub marker FasIII, H3T3P, and H3S10P. A prophase germ cell (yellow dotted outline) expressing H3T3A-GFP (A) is lack of H3T3P (B) but has abundant H3S10P (C) that co-localizes with condensed chromosome labeled by Hoechst staining (D). Two mitotic CySCs (white dotted outline) without H3T3A-GFP (A) has both H3T3P (B) and H3S10P (C) signals co-localized with condensed chromosome labeled by Hoechst staining (D). Asterisks, hub.

(E–H) Mitotic germ cells (cyan dotted outline) expressing *nos>H3-GFP* (E) as a control have both H3T3P (F) and H3S10P (G) signals co-localized with condensed chromosome labeled by Hoechst staining (H). Scale bars, 5 μm.

See also Figure S2.

Consistent with this potential function of H3T3P, immunostaining signals of H3T3P were only detectable in prophase (Figures 1N and S1D) to metaphase (Figures 1Z, and S1D), but not in late anaphase (Figure S1D) GSCs. By contrast, immunostaining using an antibody against another mitosis-enriched H3S10P (phosphorylation at serine 10 of H3) showed abundant signal throughout mitosis (Figure S1D). Furthermore, the signal from H3T3P immunostaining (Figure S1E) was enriched, but not restricted, to the centromeric region labeled with an antibody against a centromere-specific H3 variant centromere identifier (Cid) (Figure S1E). In summary, the temporal and spatial distributions of H3T3P in *Drosophila* male germ cells are comparable to what has been reported in other cell types from other systems (Caperta et al., 2008; Dai et al., 2005; Escribá and Goday, 2013; Markaki et al., 2009; Wang et al., 2010).

Expression of an H3T3A Transgene Greatly Reduces H3T3P in Mitotic Germ Cells

To understand the function of H3T3P in male germ cells, we generated fly lines with an H3-GFP transgene carrying a point mutation that converts T3 to the unphosphorylatable alanine (Ala or A, H3T3A). Expression of the H3T3A-GFP transgene in early germ cells by the *nanos-Gal4* (*nos-Gal4*) driver (Van Doren et al., 1998) greatly reduced the H3T3P signal (yellow versus white outlined cells in Figures 2A and 2B). This reduction of immunostaining signal was specific to H3T3P, as immunostaining using anti-H3S10P showed normal signals in H3T3A-expressing cells (yellow versus white outlined cells in Figure 2C). As a control, expression of the wild-type H3-GFP had no effect on either H3T3P (Figures 2E and 2F) or H3S10P (Figure 2G) signals.

Because endogenous H3 is still abundant in testes in which early germ cells are enriched with *nos*-driving H3T3A expression (Figures S2A and S2B), the absence of H3T3P signal suggests a dominant negative effect of H3T3A. The dominant negative effect of point mutations of H3 has recently been observed with several residues of histone H3 (Herz et al., 2014; Lewis et al., 2013).

Expression of H3T3A Changes the Asymmetric H3 Segregation Pattern in Mitotic GSCs

Because expression of the H3T3A provides a loss-of-function condition for H3T3P (Figures 2, S2C, and S2D), we next explored whether asymmetric histone segregation is affected in H3T3A-expressing GSCs using the dual-color labeling strategy (Figure 3A). As a control, we used a similar system with wild-type H3 and found that pre-existing H3-GFP and newly synthesized H3-mKO are asymmetrically segregated in telophase GSCs during the second mitosis after heat-shock-induced genetic switch (Figures 3B–3D), consistent with our previous report (Tran et al., 2012). By contrast, we found a dramatic shift in histone inheritance patterns from predominantly asymmetric to predominantly symmetric pattern (Figures 3H–3J), using the dual-color transgene with H3T3A (Figure 3A). Although the majority of GSCs expressing H3T3A exhibited a symmetric pattern of histone inheritance (Figures 3H–3J), we could still detect the conventional asymmetric pattern resembling that of wild-type H3 in telophase GSCs (Figures 3E–3G). Surprisingly, we also observed the inverted asymmetric pattern (Figures 3K–3M).

We reason that if pre-existing and newly synthesized histones are randomly incorporated during the first step (Figure 1B), no separation between GFP and mKO signals should be detectable
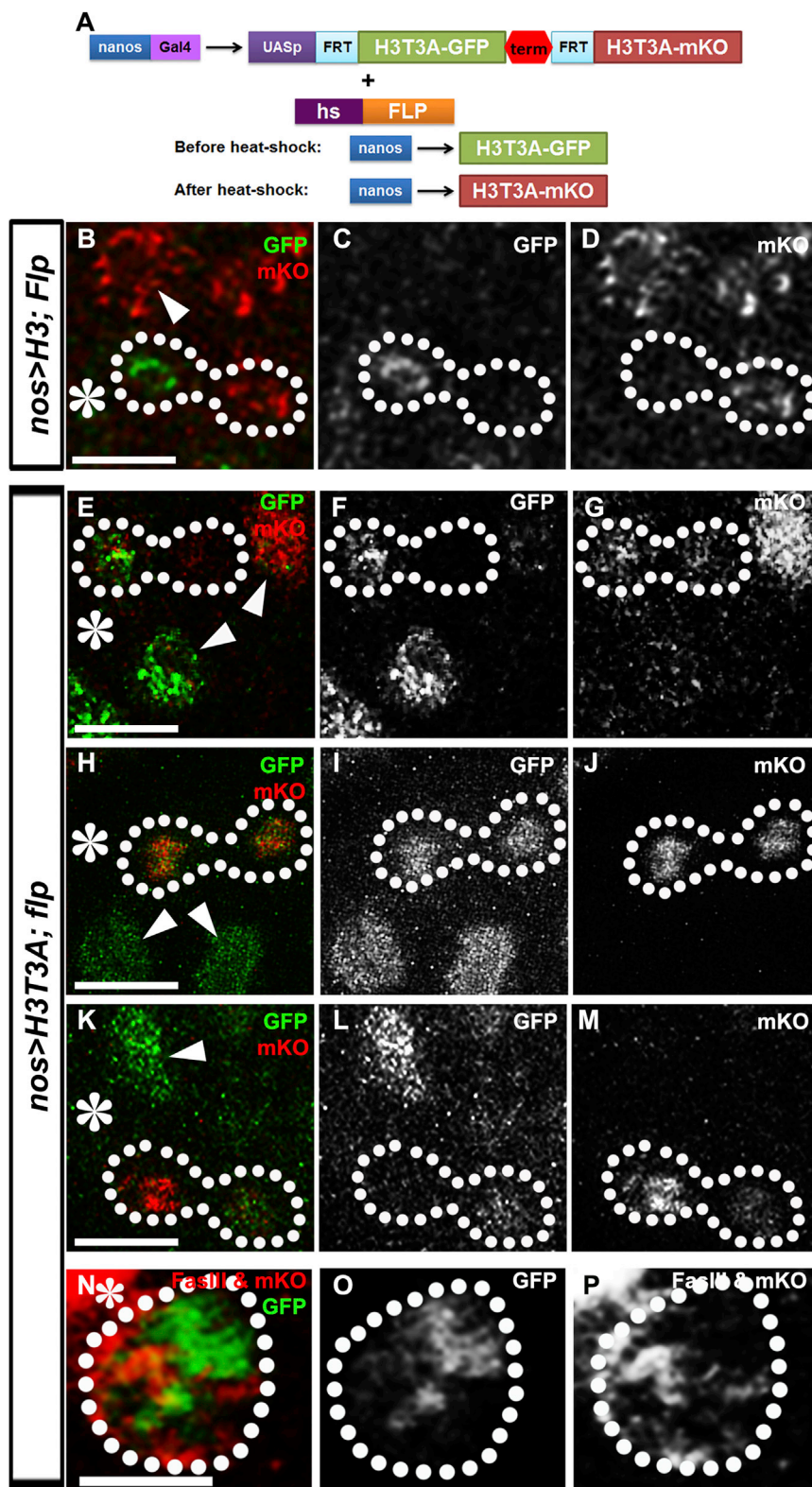
Cell

Figure 3. Expression of H3T3A Changes the Asymmetric H3 Segregation Pattern in Mitotic GSCs

(A) A schematic diagram showing the dual color-switch design that expresses pre-existing H3T3A-GFP and newly synthesized H3T3A-mKO by heat-shock treatment, as adapted from Tran et al. (2012).

(B–D) A telophase GSC expressing *nos>FRT-H3-GFP-PolyA-FRT-H3-mKO-PolyA* (*nos>H3*) during the second mitosis after heat-shock-induced genetic switch show conventional asymmetric segregation pattern.

(E–M) Telophase GSCs expressing *nos>FRT-H3T3A-GFP-PolyA-FRT-H3T3A-mKO-PolyA* (*nos>H3T3A*) during the second mitosis after heat-shock-induced genetic switch show conventional asymmetric segregation pattern (E–G), symmetric pattern (H–J), or inverted asymmetric pattern (K–M).

(N–P) A prophase GSC expressing *nos>FRT-H3T3A-GFP-PolyA-FRT-H3T3A-mKO-PolyA* (*nos>H3T3A*) during the second mitosis after heat-shock-induced genetic switch show separable GFP and mKO signals. Asterisk, hub; white dotted outline, mitotic GSCs at telophase (B–M) or prophase (N–P); arrowheads, interphase GSCs or GBs that show much less condensed nuclei. Scale bars, 5 μm.

Cell
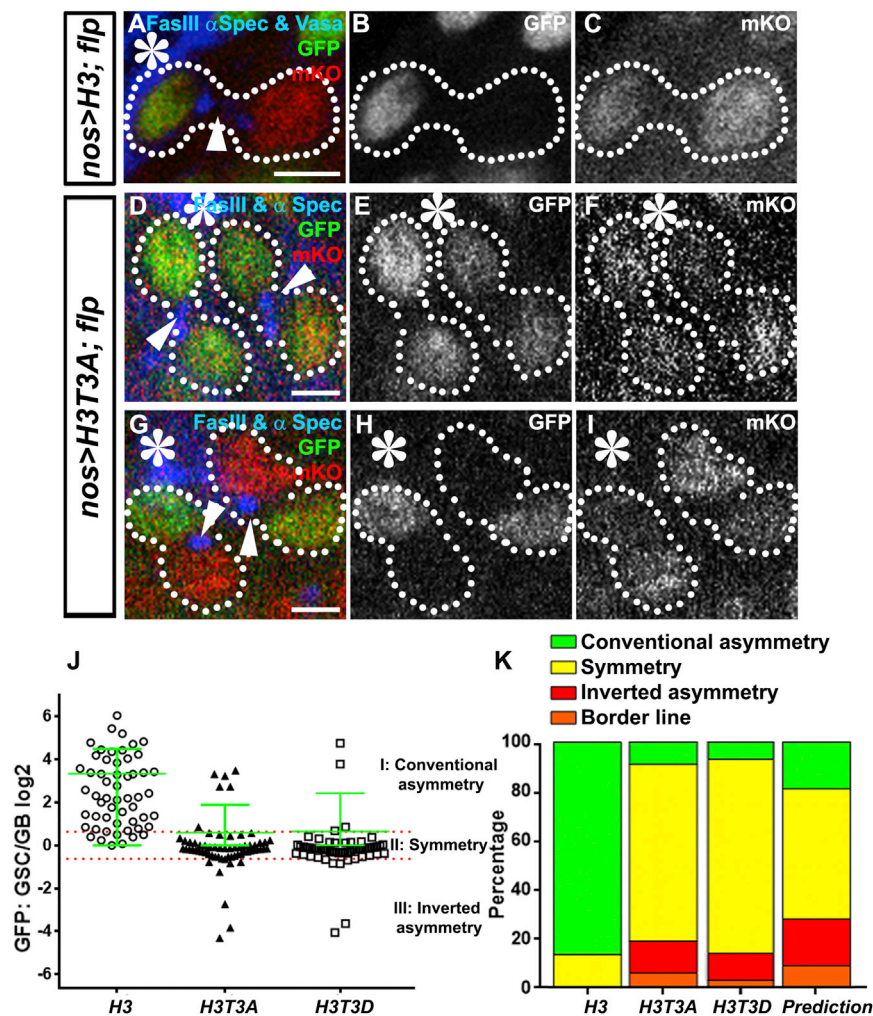
Figure 4. Expression of H3T3A or H3T3D Changes Pre-existing and Newly Synthesized H3 Distribution Patterns in Post-Mitotic GSC-GB Pairs

(A–I) Immunostaining signals using antibodies against a hub marker FasIII and spectrosome/fusome marker α-spectrin in testes from *nos>FRT-H3-GFP-PolyA-FRT-H3-mKO-PolyA* (*nos>H3*, A–C) or *nos>FRT-H3T3A-GFP-PolyA-FRT-H3T3A-mKO-PolyA* (*nos>H3T3A*, D–I) males after the second mitosis upon heat-shock-induced genetic switch. Asterisk, hub; white dotted outline, post-mitotic GSC-GB pairs; arrowheads, spectrosome structure in between GSC and GB cells. Scale bars, 5 μm.

(J) Quantification of the ratio of GFP (y axis: $\log_2$ scale) fluorescence intensity in GSC-GB pairs (see Figures S3A, S3B, and Table S1 for details): *nos>H3* (open circle, n = 55), *nos>H3T3A* (solid triangle, n = 64), and *nos>H3T3D* (open square, n = 57). Red dotted outline delineates symmetric distribution zone (see explanations below). H3 (n = 55): GSC/GB GFP ratio = 10.11 ± 1.66 (p < 10^{-4} for the ratio >1, one-tailed t test). H3T3A (n = 64): GSC/GB GFP ratio = 1.50 ± 0.28 (p > 0.05 therefore is insignificantly different from 1, two-tailed t test). H3T3D (n = 57): GSC/GB GFP ratio = 1.56 ± 0.51 (p > 0.05 therefore is insignificantly different from 1, two-tailed t test). All ratios = Avg ± SE; p value, one sample t test.

(K) Percentage of GSC-GB pairs with conventional asymmetric (GFP in GSC/GB >1.55), symmetric (GSC/GB GFP ratio between 1–1.45 and GB/GSC GFP ratio between 1–1.45), inverted asymmetric (GFP in GB/GSC >1.55), and borderline (GSC/GB GFP ratio between 1.45–1.55 and GB/GSC GFP ratio between 1.45–1.55) patterns, respectively in *nos>H3*, *nos>H3T3A*, and *nos>H3T3D* testes, as well as the predicted patterns according to randomized segregation modeling (Table S2). In *nos>H3* testes, conventional asymmetric: 87.3% (48/55); symmetric: 12.7% (7/55); no inverted asymmetric or borderline pairs. In *nos>H3T3A* testes, conventional asymmetric: 9.4% (6/64); symmetric: 71.9% (46/64); inverted asymmetric: 12.5% (8/64); borderline: 6.3% (4/64). In *nos>H3T3D* testes, conventional asymmetric: 7.0% (4/57); symmetric: 79.0% (45/57); inverted asymmetric: 10.5% (6/57); borderline: 3.5% (2/57). Predicted patterns: conventional asymmetric: 18.7% (12/64); symmetric: 53.1% (34/64); inverted asymmetric: 18.7% (12/64); borderline: 9.4% (6/64). See also Figures S3A and S3B and Tables S1 and S2.

during GSC asymmetric division. The fact that we could still identify conventional and inverted asymmetric segregation patterns in telophase GSCs (Figures 3E–3G and 3K–3M) suggests that the establishment of histone asymmetry prior to mitosis may not be affected. The observed defects in proper asymmetric segregation therefore arise upon mitotic entry when sister chromatids containing different populations of H3 need to be recognized and segregated to the appropriate daughter cell (Figure 1B, step two). Consistent with this hypothesis, separable H3T3A-GFP and H3T3A-mKO could still be detected in prophase GSCs (Figures 3N–3P and S1B), but not in a control prophase GB (Figure S1C).

Expression of H3T3A Changes H3 Distribution Patterns in Post-Mitotic GSC-GB Pairs

Since mitotic GSCs account for <2% among all GSCs (Sheng and Matunis, 2011; Yadlapalli et al., 2011; Yadlapalli and Yama-

shita, 2013), we next examined post-mitotic GSC-GB pairs derived from GSC asymmetric divisions to quantify histone inheritance patterns (Tran et al., 2012) (Experimental Procedures).

In contrast to the conventional asymmetric distribution pattern in wild-type H3-expressing GSC-GB pair (Figures 4A–4C), we observed symmetric (Figures 4D–4F), conventional asymmetric (left pair in Figures 4G–4I), and inverted asymmetric (right pair in Figures 4G–4I) distribution patterns in post-mitotic GSC-GB pairs. These data are consistent with what we have observed with mitotic GSCs (Figures 3E–3M).

Next, we quantified the percentage of each of these distribution patterns. We mainly used GFP signal to account for different patterns, for example, in Figure 4J: the conventional asymmetric patterns are in zone I, with GFP ratio in GSC/GB >1.55; the symmetric patterns are in zone II, with GFP ratio in GSC/GB <1.45 but >0.69 (i.e., GB/GSC <1.45); and the inverted asymmetric patterns are in zone III, with GFP ratio in GB/GSC >1.55.
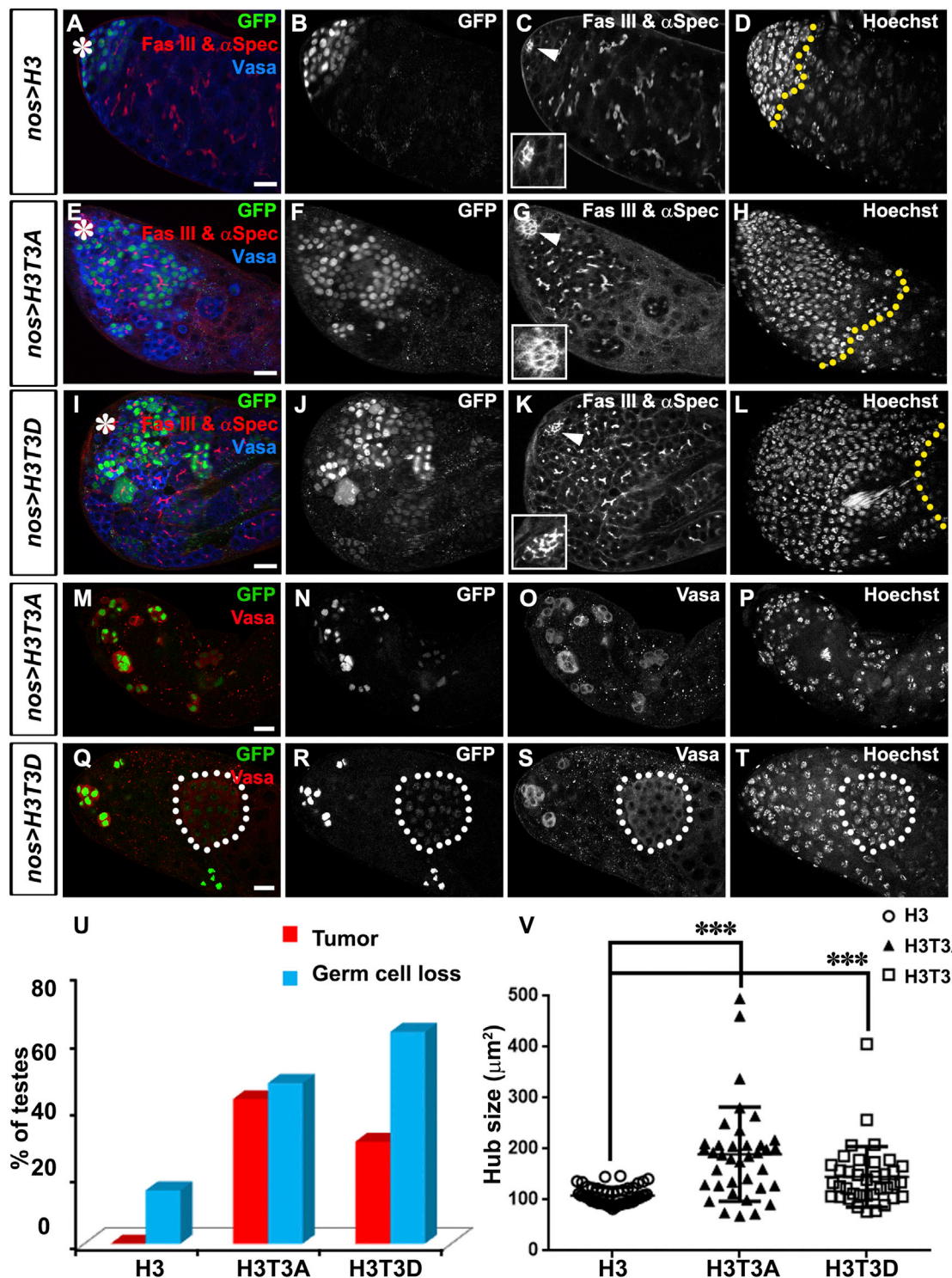
Cell

Figure 5. Both Germline and Somatic Gonadal Cells Show Defects in *nos>H3T3A* or *nos>H3T3D* Testes

(A–L) Immunostaining using antibodies against a hub marker FasIII and spectrosome/fusome marker α-spectrin in testes from *nos>H3-GFP* (A–D), *nos>H3T3A-GFP* (E–H), or *nos>H3T3D-GFP* (I–L) males 7 days after eclosion. Asterisks in (A), (E), and (I), hub; arrowheads in (C), (G), and (K) point to the hub region, which are shown with higher magnification in insets: hub size increases in *nos>H3T3A-GFP* (inset in G) or *nos>H3T3D-GFP* (inset in K) testes, but not in *nos>H3-GFP* (inset in C) testes. Early-stage germ cells, as determined by *nos*-driven GFP expression (B, F, and J), and nuclear morphology (Chen et al., 2013; Tran et al., 2000) are delineated by the yellow dotted lines in (D), (H), and (L). Scale bars, 20 μm.

(legend continued on next page)

The ~1.5-fold cutoff is based on the quantification range of symmetric H3 distribution in spermatogonial cells and symmetric H3.3 distribution in GSC-GB pairs (Tran et al., 2012, 2013). We reasoned that GFP ratio reflects the establishment of asymmetric histone distribution on sister chromatids more reliably than mKO ratio for two reasons. First, when we measured mKO fluorescence intensity in post-mitotic GSC-GB pairs, both cells are actively undergoing S phase for the next mitosis and exhibit robust incorporation of mKO-labeled newly synthesized histones (Figure 4C). Second, any histone turn-over that incorporates newly synthesized mKO-labeled histones (Deal et al., 2010; Dion et al., 2007) during processes such as transcription may not be sister chromatid-specific.

When we quantified the GFP distribution patterns in post-mitotic GSC-GB pairs in H3T3A-expressing testes (Figures 4J and S3A), we found that 71.9% (46/64) of pairs showed a symmetric pattern of inheritance (Figure 4K; Table S1). By contrast, in wild-type H3-expressing testes, 87.3% (48/55) of pairs showed an asymmetric pattern of inheritance (Figure 4K; Table S1). Moreover, in H3T3A-expressing testes, asymmetric patterns could be observed in two distinct modes at lower frequencies: 9.4% (6/64) conventional asymmetry, 12.5% (8/64) inverted asymmetry, and 6.3% (4/64) at the borderline (1.45- to 1.55-fold) between asymmetry and symmetry (Figure 4K; Table S1). Noticeably, no GSC-GB pair showed the inverted asymmetric pattern (zone III in Figure 4J) in wild-type H3-expressing testes (Figures 4J and 4K), suggesting that such a pattern is specifically induced by H3T3A-expression.

Expression of H3T3A Causes Several Germline Defects

A spectrum of cellular defects could be detected in *nos>H3T3A* testes after the level of H3T3P is effectively reduced (Figures S2C and S2D). Compared to testes expressing the wild-type H3 (Figures 5A–5D, S4A, S4D, and S5), H3T3A-expressing testes exhibited phenotypes with both germline and somatic defects (Figures 5E–5H, 5M–5P, S4B, S4E, and S5B). First, GSCs expressing the H3T3A transgene were not maintained properly. In testes without transgene or expressing H3-GFP, only germ cells with dotted spectrosome structure (de Cuevas and Spradling, 1998; Hime et al., 1996; Lin et al., 1994) were detectable next to the hub cells (Figure S4A, arrows). However, in *nos>H3T3A* testes, germ cells with branched fusome structure were detected adjacent to the hub region (arrowheads in Figure S4B), suggesting that GSCs either undergo precocious differentiation or cell death, thereby allowing more differentiated spermatogonial cysts to take their place. Quantification of these two distinct cellular structures (spectrosome versus fusome) showed a significant loss of GSCs in H3T3A-expressing testes (Figure S4C). Second, we observed a significant expansion of germline tumors carrying early-stage cellular markers, including *nos*-driven GFP expression (Figures 5E, 5F, S4E, and S5B), spectrosome structure (Figures 5E, 5G, S4E, S5B, and S5D), and condensed nuclei (Chen et al., 2013; Schulz et al., 2004; Tran et al., 2000) (Figures 5H and S5B). Interestingly, based on these cellular markers, the tumors of progenitor germ cells developed in *nos>H3T3A* testes were noticeably heterogeneous (Figure S5D). For example, some tumor cells maintained strong GFP expression (Figure S5D), a mark indicative of active *nos-Gal4* activity, and exhibited spectrosome structure (Figure S5D), suggesting that they are an early-stage GSC and/or GB cell tumor. Conversely, other tumor cells exhibited loss of GFP expression and a fusome structure (Figure S5D), suggesting that they are a later-stage spermatogonial tumor. We reason that this heterogeneity in tumor types is likely due to the heterogeneity observed in histone inheritance patterns (Figures 3 and 4). Third, the *nos>H3T3A* males had gradually decreased fertility (Figure S5C), consistent with the progression of germline defects (Figure S5B) and eventual germ cell loss (Figures 5M–5P and 5U). While the progenitor germ cell tumor phenotype was not detected in *nos>H3* (n = 19) control testes, it was observed in 42.9% of *nos>H3T3A* testes (n = 42) (Figure 5U). The germ cell loss phenotype was detected in 15.8% of *nos>H3* (n = 19) control testes but in 47.6% of *nos>H3T3A* testes (n = 42) (Figure 5U). The loss of germ cells in 15.8% of control testes is likely due to age-related effect (Boyle et al., 2007; Cheng et al., 2008; Toledano et al., 2012; Wallenfang et al., 2006). Last, *nos>H3T3A* testes (Figures 5G, inset, and S4B, yellow outline) showed a substantial hub enlargement (Figure 5V) compared to *nos>H3* testes (Figures 5C, inset, and S4A, yellow outline), most likely as a secondary defect due to GSC loss as reported previously (Dinardo et al., 2011; Gönczy and DiNardo, 1996; Monk et al., 2010; Tazuke et al., 2002). In summary, development of these germline defects in adult flies suggests that H3T3P is likely required for both GSC maintenance and proper differentiation of GB.

Expression of H3T3A in Late-Stage Germ Cells or Somatic Cells Does Not Cause Germline Tumors

The GSC loss, germline tumor and hub enlargement phenotypes in *nos>H3T3A* testes were specifically caused by expressing H3T3A in early-stage germ cells. We used a later-stage germline driver, *bam-Gal4* (Cheng et al., 2008; Eun et al., 2014; Schulz et al., 2004) (Figure 6A), to turn on the same *H3T3A* transgene in four-cell and later stage germ cells. In doing so, we were able to effectively reduce H3T3P in the more differentiated germ cells (Figure 6G). However, in this population of symmetrically dividing cells, we did not detect the phenotypes (Figures 6J–6M) we had observed in *nos>H3T3A* testes (Figures 5, S4, and S5).

In addition to GSCs, another type of adult stem cell residing in the *Drosophila* testis niche is the cyst stem cell (CySC), which, under normal conditions, is the only mitotically active somatic

(M–T) Immunostaining using a germ cell-specific anti-Vasa in testes from *nos>H3T3A-GFP* (M–P) or *nos>H3T3D-GFP* (Q–T) males. Both germ cell loss (M–T) and germline tumors (white dotted outline in Q–T) are detectable. Hoechst stains nuclei in (D), (H), (L), (P), and (T). Scale bars, 20 μm.

(U) Quantification of the percentage of testes with germline tumor and/or germ cell loss in testes expressing *nos>H3-GFP* (n = 19), *nos>H3T3A-GFP* (n = 42), or *nos>H3T3D-GFP* (n = 43).

(V) Quantification of hub size: $108 \pm 2.393 \ \mu m^2$ in *nos>H3-GFP* (n = 50) testes versus $198.5 \pm 15.22 \ \mu m^2$ in *nos>H3T3A-GFP* testes (n = 37) (***p < 10^{-4}) or $145.2 \pm 9.702 \ \mu m^2$ in *nos>H3T3D-GFP* testes (n = 37) (***p < 10^{-4}). All ratios = Avg ± SE; p value calculated by unpaired t test.
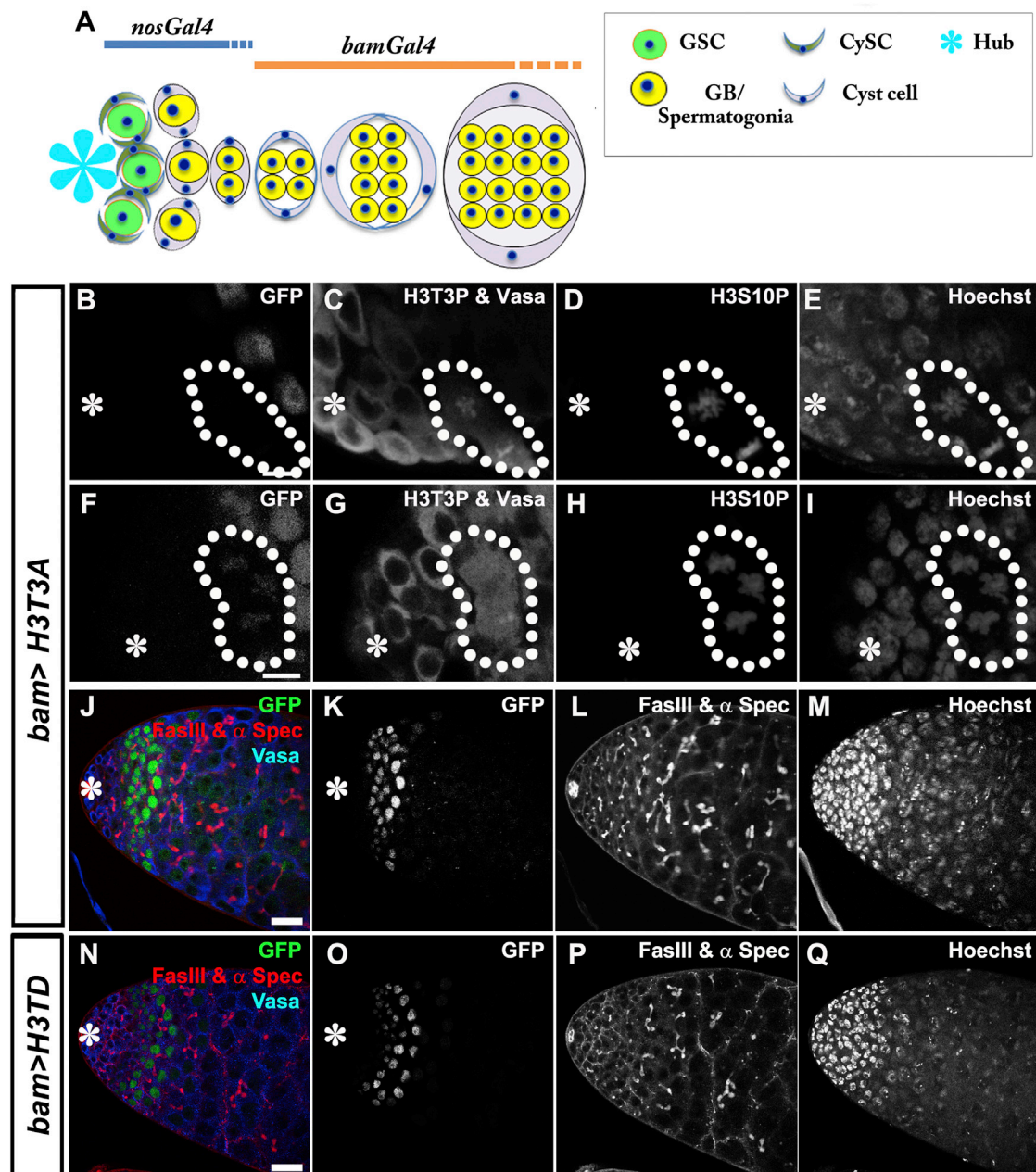
See also Figures S4 and S5.

Figure 6. Expression of H3T3A or H3T3D Using the *bam-Gal4* Driver Did Not Phenocopy Defects in *nos>H3T3A* or *nos>H3T3D* Testes

(A) A cartoon showing stage-specificity of *nos-Gal4* and *bam-Gal4* drivers: *nos-Gal4* is turned on in early-stage germline, including GSCs (Van Doren et al., 1998), while *bam-Gal4* expresses from four-cell spermatogonial cells (Cheng et al., 2008; Eun et al., 2014; Schulz et al., 2004).

(B–I) Immunostaining using antibodies against the germ cell-specific marker Vasa, H3T3P, and H3S10P in *bam>H3T3A-GFP* testes. Expression of *bam>H3T3A* greatly reduces H3T3P in later stage mitotic spermatogonial cells: a two-cell mitotic spermatogonial cyst (white dotted outline in B–E) without H3T3A-GFP (B) had detectable H3T3P (C) and H3S10P (D), both H3T3P and H3S10P overlapped with DNA signal stained with Hoechst (E). By contrast, a four-cell mitotic spermatogonial cyst (white dotted outline in F–I) with H3T3A-GFP (F) had greatly reduced H3T3P (G) but abundant H3S10P (H), the H3S10P signal overlapped with DNA signal stained with Hoechst (I). The diffusive signal in (C) and (G) came from anti-Vasa, which stains the entire mitotic germ cells because their nuclear envelopes are broken down (Yadlapalli et al., 2011; Yuan et al., 2012). Scale bars, 10 μm.

(J–Q) Immunostaining using antibodies against a hub marker FasIII, spectrosome/fusome marker α-spectrin and Vasa: tip of the testis expressing *bam>H3T3A-GFP* (J–M) or *bam>H3T3D-GFP* (N–Q). Scale bars, 20 μm. Asterisks in (B–K), (N), and (O), hub.

See also Figure S6.

gonadal cell type (Dinardo et al., 2011). When we used a somatic cell-specific *Tj-Gal4* driver (Tanentzapf et al., 2007) to express H3T3A, we found it is sufficient to reduce H3T3P signal specifically in CySCs (Figure S6A). However, no dramatic cellular defects could be detected when comparing *Tj>H3T3A* (Figure S6C) with *Tj>H3* testes (Figure S6B). In summary, these stage-specific and cell type-specific effects caused by H3T3A expression suggest that the phenotype we observed in *nos>H3T3A* testes is unlikely the result of a global perturbation of general cellular machineries.

Expression of H3T3D in Early-, but Not Late-Stage, Germ Cells Leads to Randomized H3 Inheritance and Cellular Defects

To further understand how H3T3P functions in GSCs, we expressed a different H3T3 mutant for which the T3 residue was converted to the phosphomimetic aspartic acid (D), under the hypothesis that such a mutation may disrupt the temporal order of H3T3 phosphorylation (Figures 1I–1N and S1D). Indeed, expression of H3T3D in early germ cells using a similar dual-color labeling strategy (as described for H3T3A in Figure 3A) also randomizes pre-existing H3T3D and newly synthesized H3T3D inheritance patterns (Figures 4J, 4K, and S3B; Table S1): approximately 79.0% (45/57) of GSC-GB pairs showed symmetric inheritance patterns, 7.0% (4/57) showed conventional asymmetry, and 10.5% (6/57) showed inverted asymmetry, with the remaining 3.5% (2/57) of pairs at the borderline between asymmetry and symmetry (1.45- to 1.55-fold). The randomized H3T3D inheritance patterns cannot be attributed to loss of H3T3P, as H3T3P is still detectable in H3T3D-expressing GSCs (Figure S3C). These data suggest that it is likely the timing of the H3T3 phosphorylation that is important for normal GSC activity.

In addition, both progenitor germline tumor (Figures 5I–5L and S4F) and germ cell loss (Figures 5Q–5T) phenotypes could be detected in *nos-H3T3D* testes (Figure 5U). Quantification showed significant decrease of GSCs in *nos>H3T3D* testes (6.84 ± 0.41, n = 37) compared to that of the control *nos>H3* testes (8.68 ± 0.31, n = 19; p < 0.001). Moreover, similar to the *nos>H3T3A* testes, the hub region in *nos>H3T3D* testes was also enlarged compared to the control *nos>H3* testes (Figures 5V and S4F), most likely as a secondary effect due to the loss of GSCs. By contrast, no germline tumor phenotype was found when the same transgene H3T3D-GFP was driven by the *bam-Gal4* driver (Figures 6N–6Q).

Since both reduction of H3T3P by expression of H3T3A and the mimicking of H3T3P by expression of H3T3D result in similar histone inheritance and germline defects, we hypothesize that phosphorylation of H3T3 might require a tight temporal control during GSC mitosis. Therefore, expressing either the H3T3A or the H3T3D may lead to loss of this control and similar defects in histone inheritance patterns as well as abnormal germline activity.

Differential Effects of *haspin* Gene Mutations on Germline Tumor Phenotypes in H3T3A- and H3T3D-Expressing Testes

The kinase that generates the H3T3P mark has been identified to be the Haspin protein (Dai et al., 2005). By driving a short hairpin

RNA (shRNA) (Ni et al., 2011) with the *nos-Gal4* driver to knock down *haspin*, specifically in early-stage germ cells, we were able to observe a significant decrease of H3T3P in GSCs (Figure S7A). Testes expressing *nos>haspin shRNA* showed a much greater frequency of cell death (Figures S7C and S7D) confined mainly to spermatogonial cells (Yacobi-Sharon et al., 2013), when compared to the *nos-Gal4* control (Figure S7B). Even though spermatogonial cell death was also detected in *nos>H3T3A* testes (and in *bam>H3T3A* testes), germline tumor phenotype was much more prevalent in *nos>H3T3A* testes than in *nos>haspin shRNA* testes. The similarity between *nos>H3T3A* and *nos>haspin shRNA* phenotypes is consistent with the fact that both lead to reduced H3T3 phosphorylation. The difference between *nos>H3T3A* and *nos>haspin shRNA* phenotypes suggests that the phenotypes induced by H3T3A expression are not simply a byproduct of compromising Haspin kinase activity in general. It is likely that Haspin targets some, as of yet, unknown substrates other than H3T3 in *Drosophila* GSCs. For instance, the yeast Haspin homolog has been shown to have potential roles in regulating mitotic spindle polarity (Panigada et al., 2013). It has also been reported that knockdown of Haspin in human cells (Wang et al., 2010; Yamagishi et al., 2010; reviewed by Higgins, 2010) or in *Xenopus* (Kelly et al., 2010) results in mitotic spindle defects.

To further understand potential interactions between Haspin and loss-of-H3T3P phenotypes, we first asked whether halving the level of Haspin could enhance the *nos>H3T3A* phenotype. For this, we utilized a set of permissive conditions described hereafter to create a sensitized genetic background. Due to the temperature sensitivity of the Gal4:UAS system, flies grown at lower temperature (i.e., 18°C) have been shown to have reduced levels of Gal4-driven expression (Eliazer et al., 2011). In testis samples from *nos>H3T3A* flies grown at 18°C at an earlier developmental stage (3rd instar larvae), we found that H3T3P is still abundant and cellular defects were minimal. For example, no obvious germline tumor was detected (n = 18, Figures 7A–7D). Therefore, we utilized these conditions as a permissive but sensitized genetic background. In this background, if Haspin level was halved (using a deficiency chromosome that uncovers the *haspin* gene locus; Figures 7E–7H), increased germ cell tumors could be detected (56%, n = 19, Figure 7M). These tumors were identified using a variety of morphological features, including expansion of germ cells with *nos-Gal4*-driving GFP expression (Figure 7F versus 7B), spectrosome structure (Figure 7G versus 7C), and condensed nuclei (Figure 7H versus 7D). Enhancement of the germline tumor phenotype in *nos>H3T3A* testes was also detected using a hypomorphic *haspin*[mi09386] allele (Venken et al., 2011), although with less severity (Figures 7I–7L) and lower penetrance (21%, n = 16, Figure 7M). In summary, these data showed that in *nos>H3T3A* testes the germline tumor phenotype could be enhanced by loss-of-function in *haspin* gene.

We next explored the genetic interaction between *haspin* and *nos>H3T3D* phenotype by utilizing a set of restrictive conditions—flies were grown at 18°C, shifted to 29°C as newly eclosed flies and kept at 29°C for 7 days, under which *nos>H3T3D* testes showed strong phenotype with high penetrance. We found that when Haspin level was halved using the same deficiency chromosome that uncovers the *haspin* gene
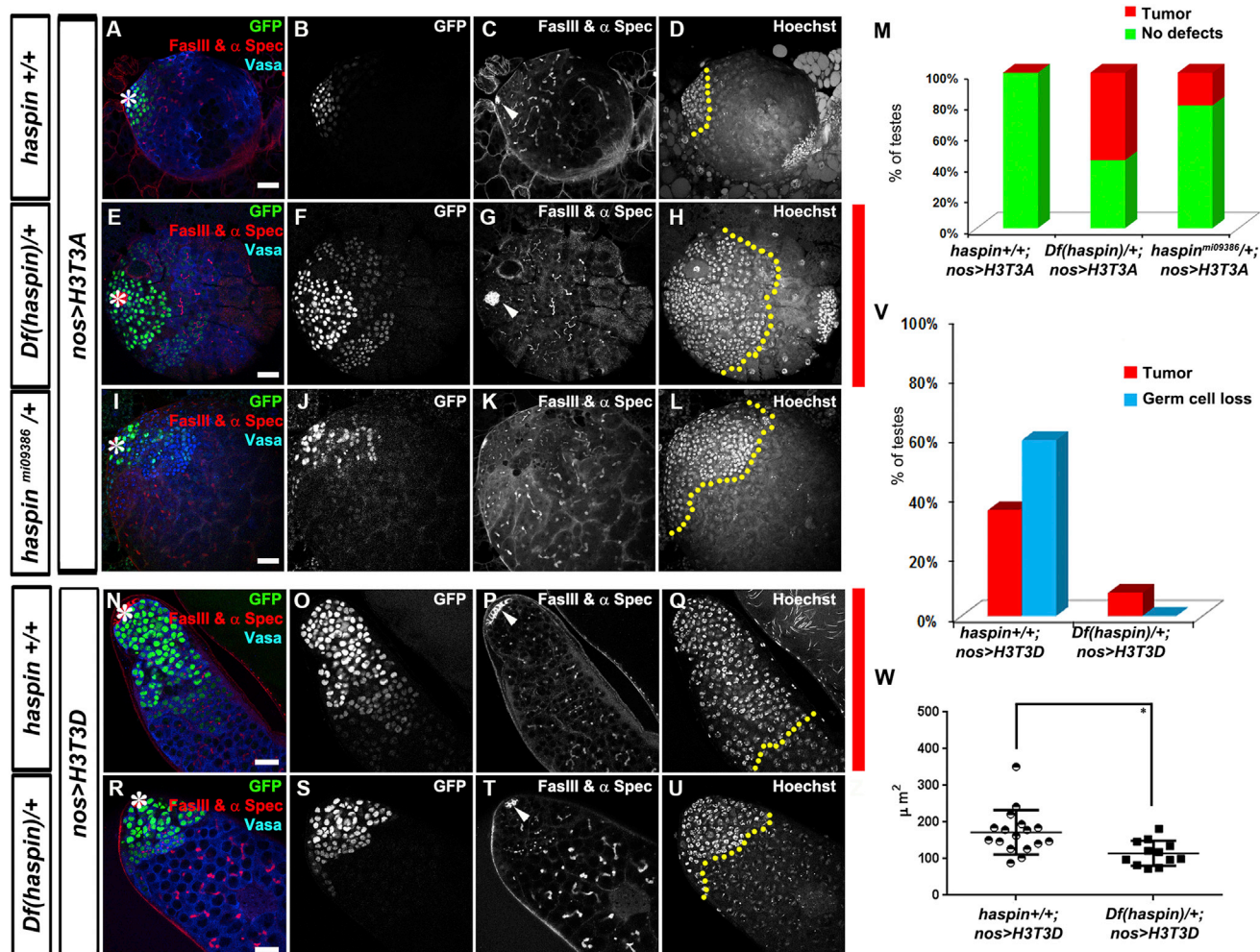
Cell

Figure 7. Genetic Interactions between *haspin* Gene Mutants and Mutations of H3T3

(A–L) Immunostaining using antibodies against a hub marker FasIII, spectrosome/fusome marker α-spectrin and germ cell marker Vasa in larval testes from *nos>H3T3A* (A–D), *Df (haspin)/+; nos>H3T3A* (E–H) or *haspin^{mi09386}/+; nos>H3T3A* (I–L) males at constant 18°C. Early-stage germline tumor is detected in testes from *Df (haspin)/+; nos>H3T3A* (F) and (H) or *haspin^{mi09386}/+; nos>H3T3A* (J) and (L) males, but not in testes from *nos>H3T3A* (B) and (D) males. Arrowhead in (G) points to enlarged hub area compared to (C).

(M) Percentage of testes that are normal or have germline tumor(s) from males of the following genotypes: *nos>H3T3A* (n = 18); *Df (haspin)/+; nos>H3T3A* (n = 16); and *haspin^{mi09386}/+; nos>H3T3A* (n = 19).

(N–U) Immunostaining using antibodies against FasIII, α-spectrin, and Vasa in testes from *nos>H3T3D* (N–Q) or *Df (haspin)/+; nos>H3T3D* (R–U) males (siblings from the same crosses) grown at 18°C, shifted to 29°C as newly eclosed flies and kept at 29°C for 7 days. Early-stage germline tumor is detected in testes from *nos>H3T3D* males (O) and (Q), but less severe in testes from *nos>H3T3D; Df (haspin)/+* males (S) and (U). Arrowhead in (P) points to enlarged hub area compared to (T). Early-stage germ cells, as determined by *nos*-driven GFP expression (B), (F), (J), (O), and (S), and nuclear morphology are delineated by the yellow dotted lines in (D), (H), (L), (Q), and (U).

(V) Percentage of testes that have germline tumor(s) or germ cell loss from males with the following genotypes: *nos>H3T3D* (n = 17) or *Df (haspin)/+; nos>H3T3D* (n = 12).

(W) Quantification of hub size: 172.2 ± 14.72 µm^2 in *nos>H3T3D* (n = 17) testes versus 115.0 ± 9.802 µm^2 in *Df (haspin)/+; nos>H3T3D* (n = 12) testes (all ratios = Avg ± SE; *$p < 0.005$, calculated by unpaired t test). Asterisks in (A), (E), (I), (N), and (R), hub. Scale bars, 20 µm.

See also Figure S7.

locus, both germline tumor and germ cell loss phenotypes in *nos>H3T3D* testes were suppressed, as indicated by lower severity (compare Figures 7N–7Q with Figures 7R–7U) and reduced penetrance (Figure 7V). Consistently, the secondary hub enlargement defect in *nos>H3T3D* testes was also suppressed (Figure 7W). These findings are reminiscent of published

studies in which expression of a phosphomimetic substrate can rescue the phenotypes of compromised kinase activity in cancer cells (Wu et al., 2010). Together, the opposite genetic interactions between *haspin* and the two H3 mutations on T3 further support the hypothesis that H3T3P needs to be tightly controlled for proper H3 inheritance and germline activity.

DISCUSSION

Here, we report that a mitosis-enriched H3T3P mark acts as a transient landmark that distinguishes sister chromatids with identical genetic code but different epigenetic information, shown as pre-existing H3-GFP and newly synthesized H3-mKO. By distinguishing sister chromatids containing different epigenetic information, H3T3P functions to allow these molecularly distinct sisters to be segregated and inherited differentially to the two daughter cells derived from one asymmetric cell division. The selective segregation of different populations of histones likely allows these two cells to assume distinct fates: self-renewal versus differentiation. Consequently, loss of proper epigenetic inheritance might lead to defects in both GSC maintenance and GB differentiation, suggesting that both cells need this active partitioning process to either "remember" or "reset" their molecular properties.

The temporal and spatial specificities of H3T3P make it a great candidate to regulate asymmetric sister chromatid segregation. First, H3T3P is only detectable from prophase to metaphase, the window of time during which the mitotic spindle actively tries to attach to chromatids through microtubule-kinetochore interactions. Second, the H3T3P signal is enriched at the peri-centromeric region, where kinetochore components robustly crosstalk with chromatin-associate factors. Third, H3T3 shows a sequential order of phosphorylation, first appearing primarily on sister chromatids enriched with pre-existing H3 and then subsequently appearing on sister chromatids enriched with newly synthesized H3 as the GSC nears metaphase. The distinct temporal patterns shown by H3T3P are unique to GSCs and would allow the mitotic machinery to differentially recognize sister chromatids bearing distinct epigenetic information; an essential step necessary for proper segregation during asymmetric GSC division. Furthermore, the tight temporal control of H3T3 phosphorylation suggests that rather than serving as an inherited epigenetic signature, H3T3P may act as transient signaling mark to allow for the proper partitioning of H3. We hypothesize that H3T3P needs to be under tight temporal control in order to ensure proper H3 inheritance and germline activity.

Our studies have shown that H3T3P is indeed subject to stringent temporal controls during mitosis. The H3T3P mark is undetectable during G2 phase. Upon entry to mitosis, sister chromatids enriched with pre-existing H3-GFP histone begin to show H3T3 phosphorylation prior to sister chromatids enriched with newly synthesized H3-mKO. As the cell continues to progress toward metaphase, H3T3P signal begins to appear on sister chromatids enriched with newly synthesized H3-mKO. Such a tight regulation of H3T3P is compromised when levels of H3T3P are altered due to the incorporation of mutant H3T3A or H3T3D. Incorporation of the H3T3A mutant results in a significant decrease in the levels of H3T3P on sister chromatids throughout mitosis, such that neither sister becomes enriched with H3T3P as the GSC progresses toward metaphase. Conversely, incorporation of the H3T3D mutant would result in seemingly elevated levels of H3T3P early in mitosis. Although H3T3A and H3T3D act in different ways, both mutations significantly disrupt the highly regulated temporal patterns associated with H3T3 phosphorylation, the result of which is randomized H3 inheritance

patterns and germ cell defects in testes expressing either H3T3A or H3T3D.

To further evaluate the extent of H3T3A and H3T3D roles in the segregation of sister chromatids enriched with different populations of H3 during mitosis (Figure 1B, step two), we modeled all possible segregation patterns in male GSCs and compared these estimates to our experimental results. To simplify our calculations, we made two important assumptions: first, we assume nucleosomal density to be even throughout the genome. This assumption allows us to infer that the overall fluorescent signal contributed by each chromosome is proportional to their respective number of DNA base pairs. Second, by quantifying pre-existing H3-GFP asymmetry in anaphase and telophase GSCs, we estimate that the establishment of H3-GFP asymmetry is ~4-fold biased, i.e., 80% on one set of sister chromatids and 20% on the other set of sister chromatids, based on quantification of GFP signal in anaphase (GFP GSC side/GB side = 4.5) and telophase (GFP GSC side/GB side = 3.8) GSCs (Tran et al., 2012). With these two simplifying assumptions, we calculate both GFP and mKO ratios among all 64 possible combinations (Table S2: 2 (for X-ch) × 2 (for Y-ch) × 4 (for 2nd ch) × 4 (for 3rd ch) = 64 combinations in total). If we define asymmetry as a greater than ~1.5-fold difference in fluorescence intensity, then based on a model of randomized sister chromatid segregation, we estimate that a symmetric pattern should appear for 53.1% (34/64) of GSC-GB pairs whereas both conventional and inverted asymmetric patterns should occur with equal frequencies and account for 18.7% (12/64) of total GSC-GB pairs. The remaining 9.4% (6/64) of GSC-GB pairs should produce histone inheritance patterns with a 1.45- to 1.55-fold difference in signal intensity (predicted ratios in Figure 4K).

This estimation is close to our experimental data in both H3T3A- and H3T3D-expressing testes (Figures 4J and 4K; Table S1). Of the 64 quantified post-mitotic GSC-GB pairs in *nos>H3T3A* testes, ~71.9% showed symmetric inheritance pattern. Conventional and inverted asymmetric patterns were detected at 9.4% and 12.5%, respectively, and 6.3% at the borderline. Similarly, of the 57 quantified post-mitotic GSC-GB pairs in *nos>H3T3D* testes, ~79.0% showed symmetric inheritance pattern. Conventional and inverted asymmetric patterns were detected at 7.0% and 10.5%, respectively with 3.5% of pairs at the borderline. Some differences between predicted ratios and our experimental data could be due to the simplified assumptions, the limited sensitivity of our measurement, and/or some coordinated chromatid segregation modes that bias the eventual read-out (Yadlapalli and Yamashita, 2013). In summary, comparison between the modeling ratios and our experimental data suggest that loss of the tight control of H3T3 phosphorylation in GSCs randomizes segregation of sister chromatids enriched with different populations of H3.

If the temporal separation in the phosphorylation of H3T3 on epigenetically distinct sister chromatids facilitates their proper segregation and inheritance during asymmetric cell division, it is likely that mutations of the Haspin kinase will also affect the temporal control of H3T3 phosphorylation. In the context of H3T3A, where the levels of H3T3P are already reduced, a further decrease in H3T3P by reducing Haspin levels should limit the GSC's ability to distinguish between sister chromatids enriched

Cell

with distinct H3. Indeed, *haspin* mutants enhance the phenotypes in *nos>H3T3A* testes. A different situation appears in the context of H3T3D where sister chromatids experience seemingly elevated levels of H3T3P at the start of mitosis. These elevated H3T3P levels may be exacerbated by the phosphorylation activity of the Haspin kinase. Therefore, it is conceivable that by halving the levels of the Haspin kinase, H3T3 phosphorylation should be reduced to a level more closely resembling wild-type. In this way, some of the temporal specificity that is lost in the H3T3D mutant is restored, resulting in suppression of the phenotypes observed in *nos>H3T3D* testes. An exciting topic for future study would be to further explore how exactly Haspin phosphorylates H3T3 in the context of chromatin and whether H3T3A and H3T3D mutations act synergistically or antagonistically in regulating asymmetric sister chromatids segregation through differential phosphorylation of a key histone residue.

It would also be interesting to understand the potential connection between asymmetric histone inheritance and another phenomenon reported by several investigators: selective DNA strand segregation (reviewed by Evano and Tajbakhsh, 2013; Rando, 2007; Tajbakhsh and Gonzalez, 2009). Recent development of the chromosome orientation fluorescence in situ hybridization (CO-FISH) technique (Falconer et al., 2010) allows study of selective chromatid segregation at single-chromosome resolution. Using this technique in mouse satellite cells, it has been demonstrated that all chromosomes are segregated in a biased manner, such that pre-existing template DNA strands are preferentially retained in the daughter cell that retains stem cell identity. Interestingly, this biased segregation becomes randomized in progenitor non-stem cells (Rocheteau et al., 2012). Using CO-FISH in *Drosophila* male GSCs, sex chromosomes have been shown to segregate in a biased manner. Remarkably, sister chromatids from homologous autosomes have been shown to co-segregate independent of any specific strand preference (Yadlapalli and Yamashita, 2013). Such findings hint at a possible epigenetic source guiding the coordinated inheritance of *Drosophila* homologous autosomes. In many cases of biased inheritance, researchers have speculated about the existence of a molecular signature that would allow the cell to recognize and segregate sister chromatids bearing differential epigenetic information (Klar, 1994, 2007; Lansdorp, 2007; Rando, 2007; Yennek and Tajbakhsh, 2013). However, the identity of such a signature has remained elusive. The work represented in this paper provides experimental evidence demonstrating that a tightly-controlled histone modification, H3T3P, is able to distinguish sister chromatids and coordinate their segregation.

Epigenetic processes play important roles in regulating stem cell identity and activity. Failure to appropriately regulate epigenetic information may lead to abnormalities in stem cell behaviors, which underlie early progress toward diseases such as cancer and tissue degeneration. Due to the crucial role that such processes play in regulating cell identity and behavior, the field has long sought to understand whether and how stem cells maintain their epigenetic memory through many cell divisions. Our results here suggest that the asymmetric segregation of pre-existing and newly synthesized H3-enriched chromosomes may function to determine distinct cell fates of GSCs versus differentiating daughter cells.

EXPERIMENTAL PROCEDURES

Heat-Shock Scheme

Flies with *UASp-FRT-H3-GFP-PolyA-FRT-H3- mKO* or *UASp-FRT-H3T3A/D-GFP-PolyA-FRT-H3T3A/D- mKO* mutant transgene were paired with *nos-Gal4* drivers. Flies were raised at 18°C throughout development until adulthood to avoid pre-flip (Tran et al., 2012). Before heat shock, 0- to 3-day-old males were transferred to vials that had been air-dried for 24 hr. Vials were submerged underneath water up to the plug in a circulating 37°C water bath for 2 hr and recovered in a 29°C incubator for indicated time before dissection and immunostaining experiments.

Temperature Shift Assay to Induce Germline Tumor in Adult Flies

Flies with *UASp-FRT-H3-GFP-PolyA-FRT-H3- mKO* or *UASp-FRT-H3T3A/D-GFP-PolyA-FRT-H3T3A/D- mKO* paired with *nos-Gal4, bam-Gal4*, or *Tj-Gal4* driver were raised at 18°C throughout development until adulthood. Newly enclosed males were collected and shifted to 29°C for indicated time before dissection and immunostaining experiments.

Immunostaining Experiments

Immunofluorescence staining was performed using standard procedures (Hime et al., 1996; Tran et al., 2012). Primary antibodies were mouse anti-α spectrin (1:50, DSHB 3A9), mouse anti-Fas III (1:50, DSHB, 7G10), mouse anti-Armadillo (1:100; DSHB, N2 7A1 clone), rabbit anti- H3T3P (1:200, Millipore 05-746R), mouse anti-H3S10P (1:2,000; Millipore, #05-806), chicken anti-CID (1:100; gift from Dr. Sylvia Erhardt, University of Heidelberg, Germany), and rabbit anti-Vasa (1:200; Santa Cruz SC-30210). Secondary antibodies were the Alexa Fluor-conjugated series (1:200; Molecular Probes). Lysotracker (Invitrogen L7528) is applied according to manufacturer recommendation. Images were taken using the Zeiss LSM 510 META or Zeiss LSM 700 Multiphoton confocal microscope with a 40× or 63× oil immersion objectives and processed using Adobe Photoshop software.

EdU Incorporation to Label GSC-GB Pair at S Phase

EdU labeling of the GSC-GB pairs at S phase was performed using Click-iT EdU Alexa Fluor 647 Imaging Kit (Life Science C10640) according to manufacturer's instructions. Dissected testes were immediately incubated in S2 medium with 100 μM EdU for 30 min at room temperature. The testes were subsequently fixed and proceed to primary antibodies (anti-FasIII, anti-α spectrin and anti-Vasa) incubation. Fluorophore conjugation to EdU was performed along manufacturer's instructions and followed by secondary antibodies incubation.

The addition of EdU facilitates recognition of the GSC-GB pairs undergoing active DNA synthesis from those without EdU, which might be arrested due to the heat-shock treatment. The cell-cycle progression is important for the incorporation and segregation of pre-existing versus newly synthesized H3.

Quantification of GFP and mKO Intensity

No antibody was added to enhance either GFP or mKO signal. Values of GFP and mKO intensity were calculated using Image J software. DAPI signal was used to determine the area of nucleus for measuring both GFP and mKO fluorescent signals, the raw reading was subsequently adjusted by subtracting fluorescence signals in the hub region used as background in both GSC and GB nuclei and compared between each other.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.002.

AUTHOR CONTRIBUTIONS

Conceptualization, J.X. and X.C.; Methodology, J.X., V.T., B.-C.C., M.W., E.B., and X.C.; Investigation, J.X., V.T., B.-C.C., M.W., C.P., and C.S.; Writing – Original Draft, J.X., M.W., and X.C.; Funding Acquisition, E.B. and X.C.; Supervision, E.B. and X.C.

Cell

REFERENCES

Betschinger, J., and Knoblich, J.A. (2004). Dare to be different: asymmetric cell division in Drosophila, C. elegans and vertebrates. Curr. Biol. *14*, R674–R685.

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. Science *330*, 612–616.

Boyle, M., Wong, C., Rocha, M., and Jones, D.L. (2007). Decline in self-renewal factors contributes to aging of the stem cell niche in the Drosophila testis. Cell Stem Cell *1*, 470–478.

Caperta, A.D., Rosa, M., Delgado, M., Karimi, R., Demidov, D., Viegas, W., and Houben, A. (2008). Distribution patterns of phosphorylated Thr 3 and Thr 32 of histone H3 in plant mitosis and meiosis. Cytogenet. Genome Res. *122*, 73–79.

Chen, H., Chen, X., and Zheng, Y. (2013). The nuclear lamina regulates germline stem cell niche organization via modulation of EGFR signaling. Cell Stem Cell *13*, 73–86.

Cheng, J., Türkel, N., Hemati, N., Fuller, M.T., Hunt, A.J., and Yamashita, Y.M. (2008). Centrosome misorientation reduces stem cell division during ageing. Nature *456*, 599–604.

Clevers, H. (2005). Stem cells, asymmetric division and cancer. Nat. Genet. *37*, 1027–1028.

Dai, J., Sultan, S., Taylor, S.S., and Higgins, J.M. (2005). The kinase haspin is required for mitotic histone H3 Thr 3 phosphorylation and normal metaphase chromosome alignment. Genes Dev. *19*, 472–488.

de Cuevas, M., and Spradling, A.C. (1998). Morphogenesis of the Drosophila fusome and its implications for oocyte specification. Development *125*, 2781–2789.

Deal, R.B., Henikoff, J.G., and Henikoff, S. (2010). Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. Science *328*, 1161–1164.

Dinardo, S., Okegbe, T., Wingert, L., Freilich, S., and Terry, N. (2011). lines and bowl affect the specification of cyst stem cells and niche cells in the Drosophila testis. Development *138*, 1687–1696.

Dion, M.F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., and Rando, O.J. (2007). Dynamics of replication-independent histone turnover in budding yeast. Science *315*, 1405–1408.

Eliazer, S., Shalaby, N.A., and Buszczak, M. (2011). Loss of lysine-specific demethylase 1 nonautonomously causes stem cell tumors in the Drosophila ovary. Proc. Natl. Acad. Sci. USA *108*, 7064–7069.

Escribá, M.C., and Goday, C. (2013). Histone H3 phosphorylation and elimination of paternal X chromosomes at early cleavages in sciarid flies. J. Cell Sci. *126*, 3214–3222.

Eun, S.H., Shi, Z., Cui, K., Zhao, K., and Chen, X. (2014). A non-cell autonomous role of E(z) to prevent germ cells from turning on a somatic cell marker. Science *343*, 1513–1516.

Evano, B., and Tajbakhsh, S. (2013). Sorting DNA with asymmetry: a new player in gene regulation? Chromosome Res. *21*, 225–242.

Falconer, E., Chavez, E.A., Henderson, A., Poon, S.S., McKinney, S., Brown, L., Huntsman, D.G., and Lansdorp, P.M. (2010). Identification of sister chromatids by DNA template strand sequences. Nature *463*, 93–97.

Gönczy, P., and DiNardo, S. (1996). The germ line regulates somatic cyst cell proliferation and fate during Drosophila spermatogenesis. Development *122*, 2437–2447.

Herz, H.M., Morgan, M., Gao, X., Jackson, J., Rickels, R., Swanson, S.K., Florens, L., Washburn, M.P., Eissenberg, J.C., and Shilatifard, A. (2014). Histone H3 lysine-to-methionine mutants as a paradigm to study chromatin signaling. Science *345*, 1065–1070.

Higgins, J.M. (2010). Haspin: a newly discovered regulator of mitotic chromosome behavior. Chromosoma *119*, 137–147.

Hime, G.R., Brill, J.A., and Fuller, M.T. (1996). Assembly of ring canals in the male germ line from structural components of the contractile ring. J. Cell Sci. *109*, 2779–2788.

Inaba, M., and Yamashita, Y.M. (2012). Asymmetric stem cell division: precision for robustness. Cell Stem Cell *11*, 461–469.

Kelly, A.E., Ghenoiu, C., Xue, J.Z., Zierhut, C., Kimura, H., and Funabiki, H. (2010). Survivin reads phosphorylated histone H3 threonine 3 to activate the mitotic kinase Aurora B. Science *330*, 235–239.

Klar, A.J. (1994). A model for specification of the left-right axis in vertebrates. Trends Genet. *10*, 392–396.

Klar, A.J. (2007). Lessons learned from studies of fission yeast mating-type switching and silencing. Annu. Rev. Genet. *41*, 213–236.

Lansdorp, P.M. (2007). Immortal strands? Give me a break. Cell *129*, 1244–1247.

Lewis, P.W., Müller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science *340*, 857–861.

Lin, H., Yue, L., and Spradling, A.C. (1994). The Drosophila fusome, a germline-specific organelle, contains membrane skeletal proteins and functions in cyst formation. Development *120*, 947–956.

Markaki, Y., Christogianni, A., Politou, A.S., and Georgatos, S.D. (2009). Phosphorylation of histone H3 at Thr3 is part of a combinatorial pattern that marks and configures mitotic chromatin. J. Cell Sci. *122*, 2809–2819.

Martin, C., and Zhang, Y. (2007). Mechanisms of epigenetic inheritance. Curr. Opin. Cell Biol. *19*, 266–272.

Monk, A.C., Siddall, N.A., Volk, T., Fraser, B., Quinn, L.M., McLaughlin, E.A., and Hime, G.R. (2010). HOW is required for stem cell maintenance in the Drosophila testis and for the onset of transit-amplifying divisions. Cell Stem Cell *6*, 348–360.

Morrison, S.J., and Kimble, J. (2006). Asymmetric and symmetric stem-cell divisions in development and cancer. Nature *441*, 1068–1074.

Ni, J.Q., Zhou, R., Czech, B., Liu, L.P., Holderbaum, L., Yang-Zhou, D., Shim, H.S., Tao, R., Handler, D., Karpowicz, P., et al. (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. Nat. Methods *8*, 405–407.

Panigada, D., Grianti, P., Nespoli, A., Rotondo, G., Castro, D.G., Quadri, R., Piatti, S., Plevani, P., and Muzi-Falconi, M. (2013). Yeast haspin kinase regulates polarity cues necessary for mitotic spindle positioning and is required to tolerate mitotic arrest. Dev. Cell *26*, 483–495.

Polioudaki, H., Markaki, Y., Kourmouli, N., Dialynas, G., Theodoropoulos, P.A., Singh, P.B., and Georgatos, S.D. (2004). Mitotic phosphorylation of histone H3 at threonine 3. FEBS Lett. *560*, 39–44.

Rando, T.A. (2007). The immortal strand hypothesis: segregation and reconstruction. Cell *129*, 1239–1243.

Rocheteau, P., Gayraud-Morel, B., Siegl-Cachedenier, I., Blasco, M.A., and Tajbakhsh, S. (2012). A subpopulation of adult skeletal muscle stem cells retains all template DNA strands after cell division. Cell *148*, 112–125.

Schulz, C., Kiger, A.A., Tazuke, S.I., Yamashita, Y.M., Pantalena-Filho, L.C., Jones, D.L., Wood, C.G., and Fuller, M.T. (2004). A misexpression screen reveals effects of bag-of-marbles and TGF beta class signaling on the Drosophila male germ-line stem cell lineage. Genetics *167*, 707–723.

Cell

Sheng, X.R., and Matunis, E. (2011). Live imaging of the Drosophila spermatogonial stem cell niche reveals novel mechanisms regulating germline stem cell output. Development *138*, 3367–3376.

Tajbakhsh, S., and Gonzalez, C. (2009). Biased segregation of DNA and centrosomes: moving together or drifting apart? Nat. Rev. Mol. Cell Biol. *10*, 804–810.

Tanentzapf, G., Devenport, D., Godt, D., and Brown, N.H. (2007). Integrin-dependent anchoring of a stem-cell niche. Nat. Cell Biol. *9*, 1413–1418.

Tazuke, S.I., Schulz, C., Gilboa, L., Fogarty, M., Mahowald, A.P., Guichet, A., Ephrussi, A., Wood, C.G., Lehmann, R., and Fuller, M.T. (2002). A germline-specific gap junction protein required for survival of differentiating early germ cells. Development *129*, 2529–2539.

Toledano, H., D'Alterio, C., Czech, B., Levine, E., and Jones, D.L. (2012). The let-7-Imp axis regulates ageing of the Drosophila testis stem-cell niche. Nature *485*, 605–610.

Tran, J., Brenner, T.J., and DiNardo, S. (2000). Somatic control over the germ-line stem cell lineage during Drosophila spermatogenesis. Nature *407*, 754–757.

Tran, V., Lim, C., Xie, J., and Chen, X. (2012). Asymmetric division of Drosophila male germline stem cell shows asymmetric histone distribution. Science *338*, 679–682.

Tran, V., Feng, L., and Chen, X. (2013). Asymmetric distribution of histones during Drosophila male germline stem cell asymmetric divisions. Chromosome Res. *21*, 255–269.

Van Doren, M., Williamson, A.L., and Lehmann, R. (1998). Regulation of zygotic gene expression in Drosophila primordial germ cells. Curr. Biol. *8*, 243–246.

Venken, K.J., Schulze, K.L., Haelterman, N.A., Pan, H., He, Y., Evans-Holm, M., Carlson, J.W., Levis, R.W., Spradling, A.C., Hoskins, R.A., and Bellen,

H.J. (2011). MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nat. Methods *8*, 737–743.

Wallenfang, M.R., Nayak, R., and DiNardo, S. (2006). Dynamics of the male germline stem cell population during aging of Drosophila melanogaster. Aging Cell *5*, 297–304.

Wang, F., Dai, J., Daum, J.R., Niedzialkowska, E., Banerjee, B., Stukenberg, P.T., Gorbsky, G.J., and Higgins, J.M. (2010). Histone H3 Thr-3 phosphorylation by Haspin positions Aurora B at centromeres in mitosis. Science *330*, 231–235.

Wu, Q., Sahasrabudhe, R.M., Luo, L.Z., Lewis, D.W., Gollin, S.M., and Saunders, W.S. (2010). Deficiency in myosin light-chain phosphorylation causes cytokinesis failure and multipolarity in cancer cells. Oncogene *29*, 4183–4193.

Yacobi-Sharon, K., Namdar, Y., and Arama, E. (2013). Alternative germ cell death pathway in Drosophila involves HtrA2/Omi, lysosomes, and a caspase-9 counterpart. Dev. Cell *25*, 29–42.

Yadlapalli, S., and Yamashita, Y.M. (2013). Chromosome-specific nonrandom sister chromatid segregation during stem-cell division. Nature *498*, 251–254.

Yadlapalli, S., Cheng, J., and Yamashita, Y.M. (2011). Drosophila male germline stem cells do not asymmetrically segregate chromosome strands. J. Cell Sci. *124*, 933–939.

Yamagishi, Y., Honda, T., Tanno, Y., and Watanabe, Y. (2010). Two histone marks establish the inner centromere and chromosome bi-orientation. Science *330*, 239–243.

Yennek, S., and Tajbakhsh, S. (2013). DNA asymmetry and cell fate regulation in stem cells. Semin. Cell Dev. Biol. *24*, 627–642.

Yuan, H., Chiang, C.Y., Cheng, J., Salzmann, V., and Yamashita, Y.M. (2012). Regulation of cyclin A localization downstream of Par-1 function is critical for the centrosome orientation checkpoint in Drosophila male germline stem cells. Dev. Biol. *361*, 57–67.

Cell

Stable Chromosome Condensation Revealed by Chromosome Conformation Capture

Graphical Abstract
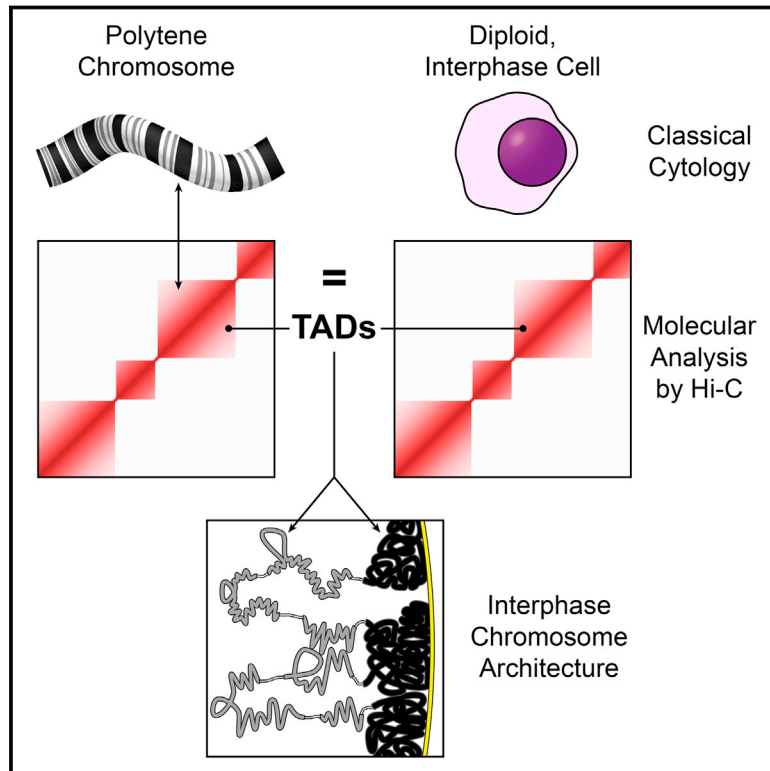
Authors
Kyle P. Eagen, Tom A. Hartl, Roger D. Kornberg

Correspondence
kornberg@stanford.edu

In Brief

Analysis of polytene bands, which are shown to correspond to topologically associating domains in interphase nuclei, reveals two stable forms of folded chromatin within euchromatic regions of diploid cells that are distinct from more highly structured heterochromatin.

Highlights

- Hi-C of polytene chromosomes reveals an equivalence of polytene bands with TADs

- TADs are conserved between polytene and diploid cells

- Fully extended and up to 10-fold compacted fibers constitute euchromatin

- Up to 30-fold compacted fibers represent heterochromatin of the nuclear periphery

Accession Numbers

GSE72512

CrossMark

CellPress

Article

Stable Chromosome Condensation Revealed by Chromosome Conformation Capture

Kyle P. Eagen,[1] Tom A. Hartl,[2,3] and Roger D. Kornberg[1,*]
[1]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
[2]Departments of Developmental Biology, Genetics, and Bioengineering, Stanford University School of Medicine, Stanford, CA 94305, USA
[3]Present address: Perlstein Lab, QB3@953, 953 Indiana Street, San Francisco, CA 94107, USA
*Correspondence: kornberg@stanford.edu
http://dx.doi.org/10.1016/j.cell.2015.10.026

SUMMARY

Chemical cross-linking and DNA sequencing have revealed regions of intra-chromosomal interaction, referred to as topologically associating domains (TADs), interspersed with regions of little or no interaction, in interphase nuclei. We find that TADs and the regions between them correspond with the bands and interbands of polytene chromosomes of *Drosophila*. We further establish the conservation of TADs between polytene and diploid cells of *Drosophila*. From direct measurements on light micrographs of polytene chromosomes, we then deduce the states of chromatin folding in the diploid cell nucleus. Two states of folding, fully extended fibers containing regulatory regions and promoters, and fibers condensed up to 10-fold containing coding regions of active genes, constitute the euchromatin of the nuclear interior. Chromatin fibers condensed up to 30-fold, containing coding regions of inactive genes, represent the heterochromatin of the nuclear periphery. A convergence of molecular analysis with direct observation thus reveals the architecture of interphase chromosomes.

INTRODUCTION

The basis of DNA folding and compaction in nuclei and chromosomes is one of the great mysteries of biology. How are two meters of DNA packaged in an interphase nucleus ~10 μm in diameter? At the molecular level, folding begins with wrapping of DNA around histones in the nucleosome (Kornberg, 1974; Luger et al., 1997). At a cytological level, chromatin appears as darkly staining heterochromatin, most abundant at the periphery of the interphase nucleus, and lightly staining euchromatin, in the nuclear interior. The folding of chromatin fibers in heterochromatin and their organization in euchromatin have not been determined. We have gained insight into chromosome condensation from molecular analysis of polytene chromosomes of *Drosophila*.

Polytene chromosomes occur in cells of dipteran larvae that undergo as many as ten rounds of DNA replication without division, while retaining close alignment of sister chromatids and pairing of homologous chromosomes (Urata et al., 1995). Poly-

tene chromosomes are visible by light microscopy (Agard and Sedat, 1983; Urata et al., 1995) and display an alternation of dense bands with less dense interbands (Balbiani, 1881; Flemming, 1882), which has been exploited for genetic analysis (Bridges, 1935). The increased density in bands reflects elevated DNA content and a stable state of chromatin condensation (Beermann, 1972). Two types of bands have been distinguished, loosely compacted gray bands and dense, tightly compacted intercalary heterochromatin (IH) bands (Vatolina et al., 2011; Zhimulev et al., 2014). Median expression levels of genes within gray bands are 27 times greater than that of genes within IH bands (Zhimulev et al., 2014).

Chromatin folding has been detected at the molecular level, by "chromosome conformation capture," in which the chemical cross-linking of chromosomal material is followed by fragmentation, ligation, and DNA sequence analysis (Dekker et al., 2002). Extension of this approach by high-throughput sequencing (Hi-C) revealed, at a resolution of 1 Mb, a genome-wide interaction network of the human genome (Lieberman-Aiden et al., 2009). Computational analysis of the Hi-C results identified two sets of chromosomal regions, termed compartments, within which very distant interactions occur more frequently than expected for the random coil configuration of a polymer. The two compartments correlate with regions of transcriptional activity and inactivity.

Increased sequencing depth revealed finer details of chromosome folding, at a resolution of 40 kb or better. So-called topologically associating domains (TADs) (Nora et al., 2012), also referred to as "physical domains" (Hou et al., 2012; Sexton et al., 2012) or "topological domains" (Dixon et al., 2012), in which nucleotide sequences far apart along the DNA come in close proximity to one another are observed at a length scale of a few Mb pairs or less. Hi-C also revealed loops, frequently bridging between enhancers and promoters, that correlate with gene activation. Loops are apparently transient, >250 nm apart in three-quarters of a cell population (Rao et al., 2014), and although important for gene regulation, are unlikely to provide a structural basis for heterochromatin and euchromatin. TADs, however, represent a consistent feature among cell types (Dixon et al., 2012, 2015) and may therefore relate to stably folded states of chromatin. In studies performed to date, however, TADs have only been revealed by DNA sequencing and not directly related to chromosome condensation. A hierarchical relationship between TADs and compartments has been suggested by computational models (Gibcus and Dekker, 2013;
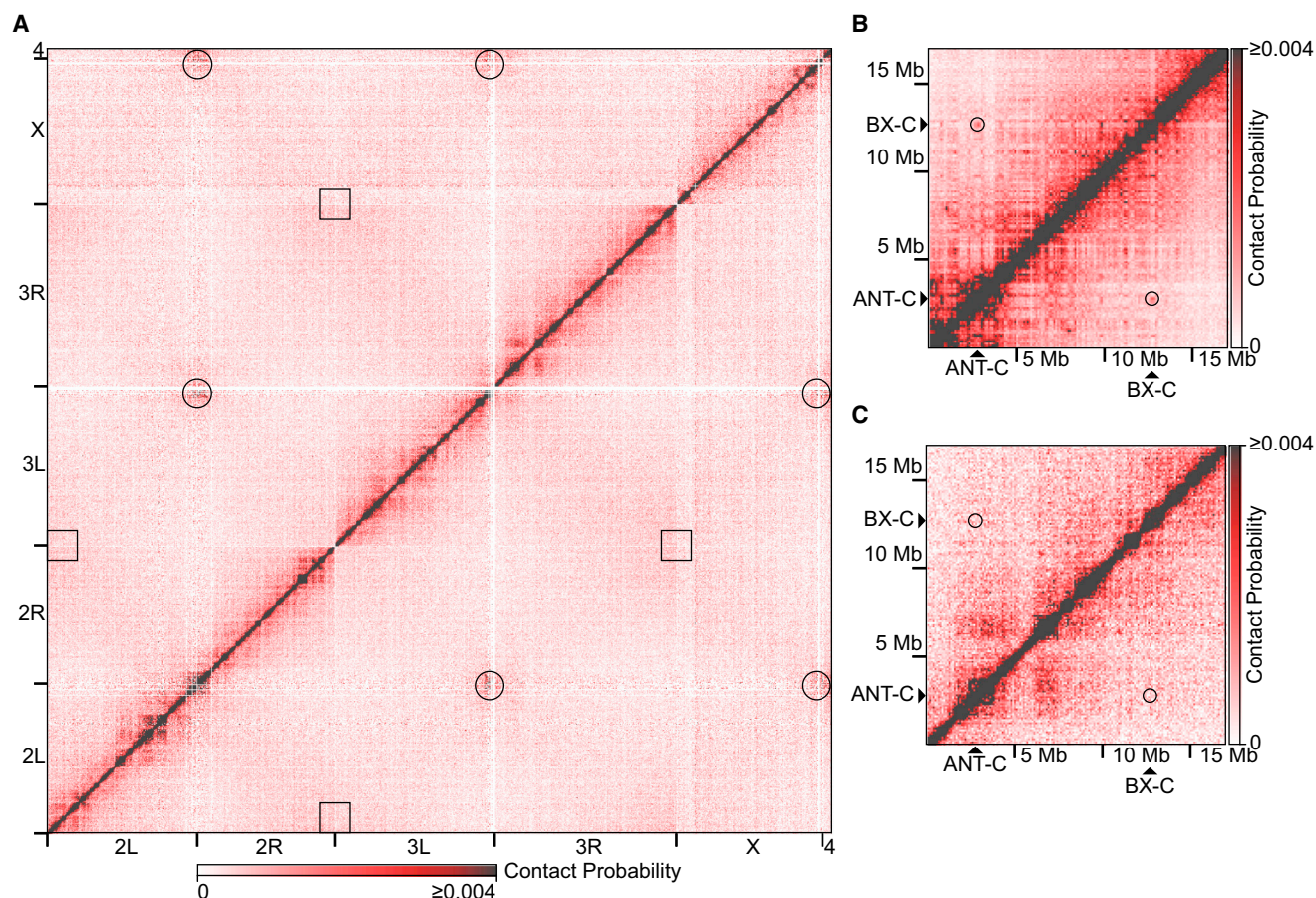
Figure 1. Lack of Regular, Long-Range Contacts in Polytene Chromosomes
(A) Genome-wide Hi-C heatmap from polytene cells. Black circles and squares represent where centromeres and telomeres intersect, respectively.
(B) Hi-C heatmap from *Drosophila* embryos (Sexton et al., 2012) of a 17 Mb region of chromosome 3R encompassing the ANT-C and BX-C loci. Black circles represent where the ANT-C and BX-C loci intersect.
(C) Hi-C heatmap from polytene cells of a 17 Mb region of chromosome 3R encompassing the ANT-C and BX-C loci. Black circles represent where the ANT-C and BX-C loci intersect.
Heatmaps were normalized and divided into 100 kb bins. See also Figures S1 and S2.

Sexton et al., 2012). We report here on the direct visualization of TADs in polytene chromosomes of *Drosophila*, from which the relationship between TADs and compartments, as well as structural correlates of chromatin condensation in the interphase nucleus, may be derived.

RESULTS

Hi-C on *Drosophila* Polytene Chromosomes
We performed Hi-C on the salivary glands of wandering third instar larvae of *Drosophila melanogaster* (Figures 1A, 1C, S1, and S2). The limited amount of material available from manually dissected, primary polytene tissue necessitated a Hi-C approach with improved signal-to-noise and limited the resolution of our analysis to 15 kb (Figure S2; Supplemental Experimental Procedures). To make reliable comparisons between the polytene Hi-C data, other Hi-C datasets, and cytological observations, we only considered chromosomal features of 75 kb

or larger. The highly underreplicated and repetitive nature of pericentromeric heterochromatin also precluded reliable Hi-C analysis in these regions. Differences in copy number across the arms of polytene chromosomes due to incomplete DNA replication do not affect Hi-C results (Figure S3; Supplemental Experimental Procedures).

Our genome-wide, polytene Hi-C heatmaps revealed interactions between centromeres, but a lack of contact between telomeres (Figure 1A). These results are consistent with three-dimensional reconstructions of polytene chromosomes by light microscopy, in which centromeres were observed to cluster at the chromocenter at one pole of the nucleus with telomeres spread in the opposite hemisphere in a Rabl orientation (Hochstrasser et al., 1986). Centromeres are also clustered together in diploid *Drosophila* Kc167 cultured cells (Hou et al., 2012) and late-stage embryos (Sexton et al., 2012). Telomeres do not interact with one another in diploid cells (Hou et al., 2012), but do interact in embryos (Sexton et al., 2012).

A

Bands Boxed

TADs Boxed

B

C

D

■ Polytene TADs ■ Between Polytene TADs

E
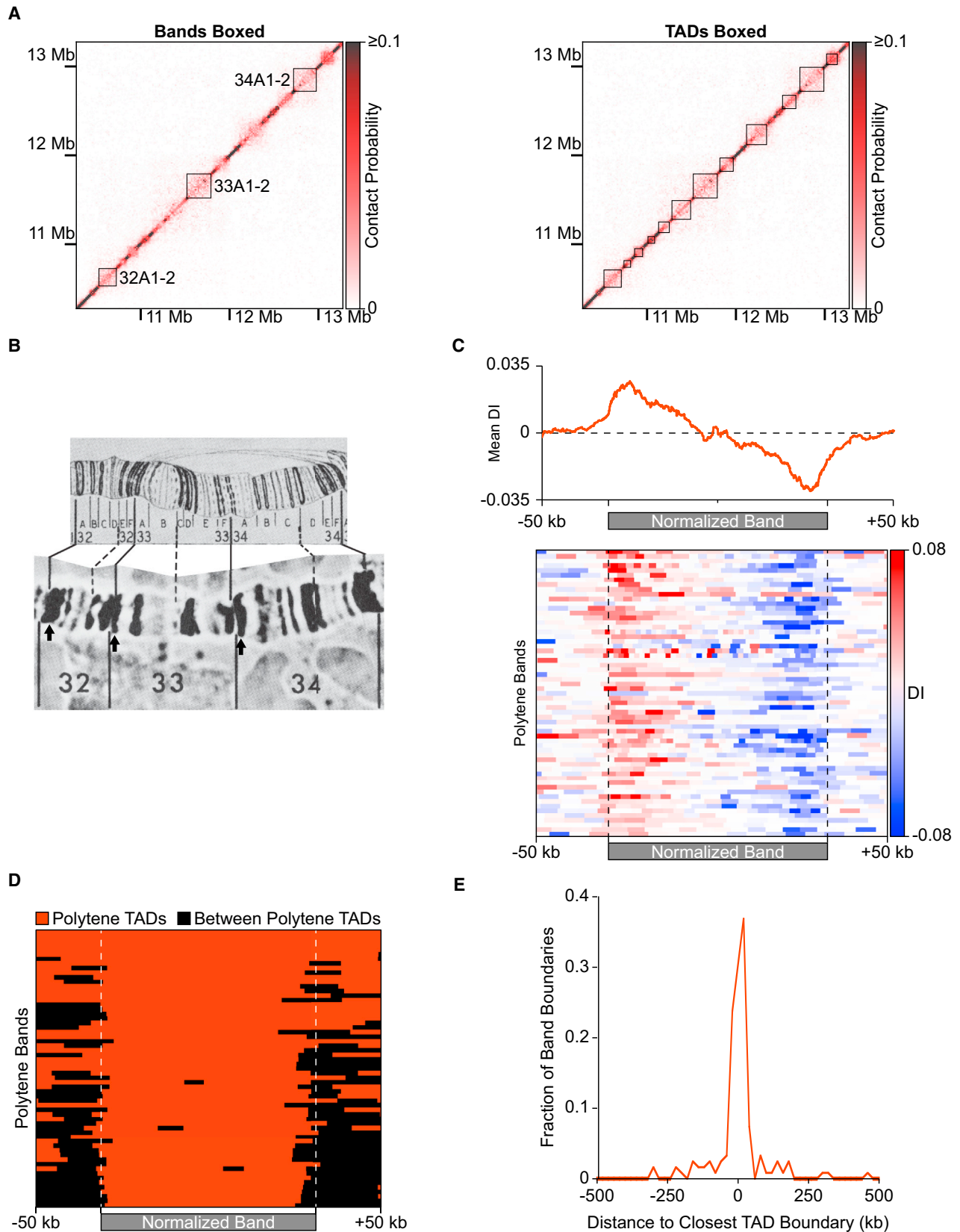
(legend on next page)

Distant or long-range interactions (separated by >1 Mb), such as those previously observed between the ANT-C and BX-C loci in Hi-C heatmaps of late-stage (mixed cell-type) *Drosophila* embryos (Sexton et al., 2012) (Figure 1B) were not detected in our polytene Hi-C heatmaps (Figure 1C). The lack of long-range interactions in our heatmaps was consistent with previously described light microscopy of polytene chromosomes in situ, which revealed only regular, short-range contacts and no reproducible, long-range interactions (Hochstrasser et al., 1986). Occasionally polytene bands interact with one another, resulting in ectopic pairing. The frequency of cells exhibiting ectopic pairing is low (Zhimulev et al., 1982) and complete mixing of bands is not observed, which explains why, at present, such pairing falls below our limit of detection by Hi-C.

We examined the proportion of paired-end reads that mapped to the same restriction fragment and pointed in the same direction, reflecting interactions between paired and aligned chromatids or homologous chromosomes (Sexton et al., 2012). A similar proportion of paired-end reads fell into this class for our polytene chromosome dataset (0.0328%) and the reported embryonic dataset (0.0553%). These values are much less than the proportion of paired-end reads representing intramolecular self-ligation (cyclization) events (0.117% and 0.132% in the polytene and embryo datasets, respectively) indicating that reads dues to interactions between paired and aligned chromatids or homologous chromosomes are very infrequent. A higher proportion of reads in the reported diploid chromosome (Kc167 cell line) dataset fall in this class (22.5% compared to 8.35% of paired-end reads representing intramolecular self-ligation events), but the chromatin was fragmented to a median size of 2,377 bp, compared with 193 bp in the polytene and embryonic experiments. Homologous chromosome pairing is prevalent in many *Drosophila* primary tissues and cultured cells (Fung et al., 1998; Williams et al., 2007), but the analysis of paired-end reads indicates that paired homologs or chromatids are not perfectly aligned at the level of a few hundred base pairs, even in polytene chromosomes, and the degree of alignment in polytene chromosomes is approximately the same as that in late-stage embryos.

Equivalence of Polytene Bands with TADs
Our polytene Hi-C heatmaps showed the presence of TADs, regions of high self-interaction, visible as boxes centered on the diagonal (Figures 2A, 5A, and S3). Our data revealed 346 polytene TADs. The median TAD size was 165 kb; the mean TAD size was 195 kb (Figure S3; Table S1). Superimposing the Hi-C heatmap onto the locations of polytene bands for which reliable DNA sequence coordinates have previously been determined by fluorescence in situ hybridization (FISH) and non-histone protein localization (Belyaeva et al., 2012; Vatolina et al., 2011) demonstrated a correspondence of TADs with bands (Figures 2A, 2B, and 5A).

For a quantitative assessment of the interaction pattern within polytene bands, we computed the genome-wide directionality index, a measure of the degree of bias of a locus for interaction with downstream (positive directionality index) or upstream loci (negative directionality index) (Dixon et al., 2012). Characteristically, the directionality index of a TAD starts positive, goes to zero at the middle of the TAD, and continues toward negative values at the end of a TAD (Dixon et al., 2012). Polytene bands exhibit this same trend, supporting their identification as TADs (Figure 2C).

The concordance between polytene bands and TADs was observed across all of the large chromosome arms (Figure 2D). At least 95% of polytene bands corresponded to uninterrupted TADs. The overlap between polytene bands and TADs was far greater than expected on a random basis (Z score = 10.5, p = 3.01×10^{-25}; Figure S4B; Supplemental Experimental Procedures). Differences were observed almost exclusively at band boundaries. Mismatches between polytene bands and TADs were far less than expected on a random basis (Z score = -10.5, p = 2.78×10^{-25}; Figure S4B; Supplemental Experimental Procedures). The majority (57.4%) of band boundaries were located within 20 kb of TAD boundaries (Figure 2E). Discrepancies at the boundaries were likely due to the limited resolution of the Hi-C experiment, limited accuracy of DNA sequence coordinates for polytene band borders, or both. Near identity of boundaries does not necessarily indicate equivalence of entire domains, but quantitative, unbiased examination of entire polytene bands revealed far greater equivalence with polytene TADs than expected on a random basis (Figure S5B; Supplemental Experimental Procedures).

Polytene Puffs Are Not TADs
High transcriptional output of some genes at certain stages of development causes the conversion of polytene bands to puffs, regions of decondensed chromatin with a diameter wider than the rest of the chromosome (Ashburner, 1967). Loci that are known to convert from bands to puffs at the wandering third

Figure 2. Polytene Bands Are TADs
(A) Normalized Hi-C heatmap (15 kb bins) of a 3 Mb region of polytene chromosome 2L. Left: areas bounded by black boxes represent locations of polytene bands for which reliable DNA sequence coordinates are available (Belyaeva et al., 2012; Vatolina et al., 2011). Right: areas bounded by black boxes represent TADs.
(B) Photographic image (bottom) of the region of polytene chromosome 2L from (A) and the same region of Bridges's chromosome map (top). Arrows indicate bands represented by black boxes in (A). The DNA sequence coordinates of other bands in this region are not known and therefore cannot be compared with the Hi-C data. Adapted from Lefevre (1976). To prepare the panel, the original photographic image from Lefevre was digitally scanned and then cropped to display the region of the chromosome corresponding to that of the Hi-C analysis in (A). Black arrows were then superimposed onto the cropped image using Adobe Illustrator to mark the polytene bands represented by black boxes in (A).
(C) Mean directionality index (DI) of polytene bands (upper panel; n = 61) and heatmap of the directionality index of each band along its length (lower panel). Bands were normalized to the same length and 50 kb of flanking DNA is shown next to each normalized band.
(D) Heatmap of the agreement between polytene bands (n = 61) and TADs. Each band is represented by a row. Bands were normalized to the same length and 50 kb of flanking DNA is shown next to each normalized band. Orange segments overlap with polytene TADs, black segments overlap with regions between TADs.
(E) Fraction of band boundaries (n = 122) at the distance indicated on the abscissa from the closest TAD boundary (calculated in 20 kb windows).
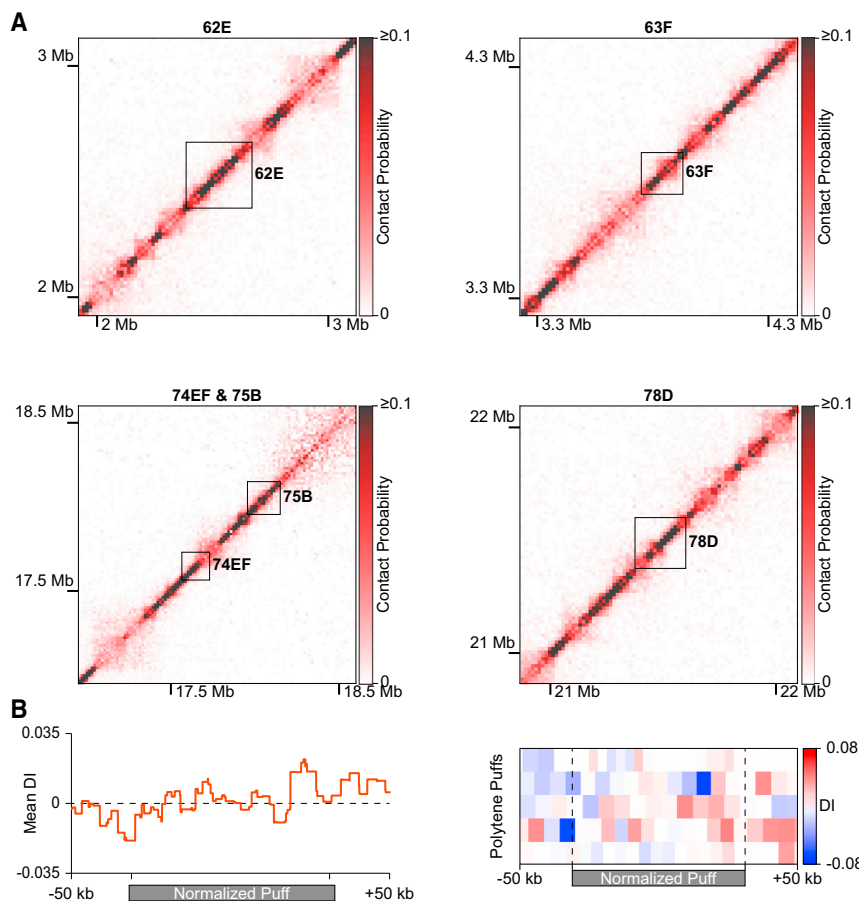See also Figures S3, S4, and S5 and Table S1.

Figure 3. Hi-C of Polytene Puffs

(A) Normalized Hi-C heatmaps (15 kb bins) of puff stage five to eight polytene puffs on chromosome 3L. Areas bounded by black boxes correspond to the indicated puff.

(B) Mean directionality index (DI) of polytene puffs (left) and heatmap of the directionality index of each puff along its length (right). Puffs were normalized to the same length and 50 kb of flanking DNA is shown next to each normalized puff.

pected (p = 1.80 × 10⁻⁶, binomial test). FISH thus confirms our Hi-C results and our identification of TADs with polytene bands.

Conservation of TADs between Polytene and Diploid Cells

Hi-C was previously performed on *D. melanogaster* diploid, Kc167 cultured cells (Hou et al., 2012), and late-stage embryos (Sexton et al., 2012). Both studies revealed TADs as an organizational feature of the *Drosophila* genome. The Hi-C heatmap from diploid cells and our Hi-C heatmap from polytene cells were closely similar (Figures 5A and 5B) and highly correlated (genome-wide Pearson's r = 0.793, p < 2.2 × 10⁻¹⁶). Where comparisons could be made, diploid TADs could also be seen to correspond to polytene bands (Figures 5B and 5C). The overlap between polytene TADs and diploid TADs (Figure 5D) was far greater than expected on a random basis (Z score = 20.5, p = 3.92 × 10⁻⁹²; Figure S4B; Supplemental Experimental Procedures). Similarly, mismatches between polytene bands and TADs were far less than expected on a random basis (Z score = −20.6, p = 6.89 × 10⁻⁹³; Figure S4B; Supplemental Experimental Procedures). Approximately 50% of polytene TAD boundaries were located within 40 kb of diploid TAD boundaries (Figure 5E) and quantitative assessment of entire polytene TADs revealed far greater equivalence with Kc cell TADs than expected on a random basis (Figure S5C; Supplemental Experimental Procedures). We observed a similar correspondence in the Hi-C heatmaps between polytene cells and late-stage embryos (Figure S6; genome-wide Pearson's r = 0.793, p < 2.2 × 10⁻¹⁶). TADs from polytene cells and *Drosophila* embryos significantly overlapped (Z score = 15.2, p = 2.10 × 10⁻⁵¹; Figure S4B; Supplemental Experimental Procedures) and exhibited far fewer mismatches than expected on a random basis (Z score = −15.3, p = 8.19 × 10⁻⁵²; Figure S4B; Supplemental Experimental Procedures). The agreement was even more remarkable considering that late-stage embryos contain a mixture of cell-types, unlike the more homogenous composition of the salivary gland. Evidently, most TADs are conserved in their central regions across a range of cells types, while a fraction of TAD boundaries may exhibit some cell-type specificity.
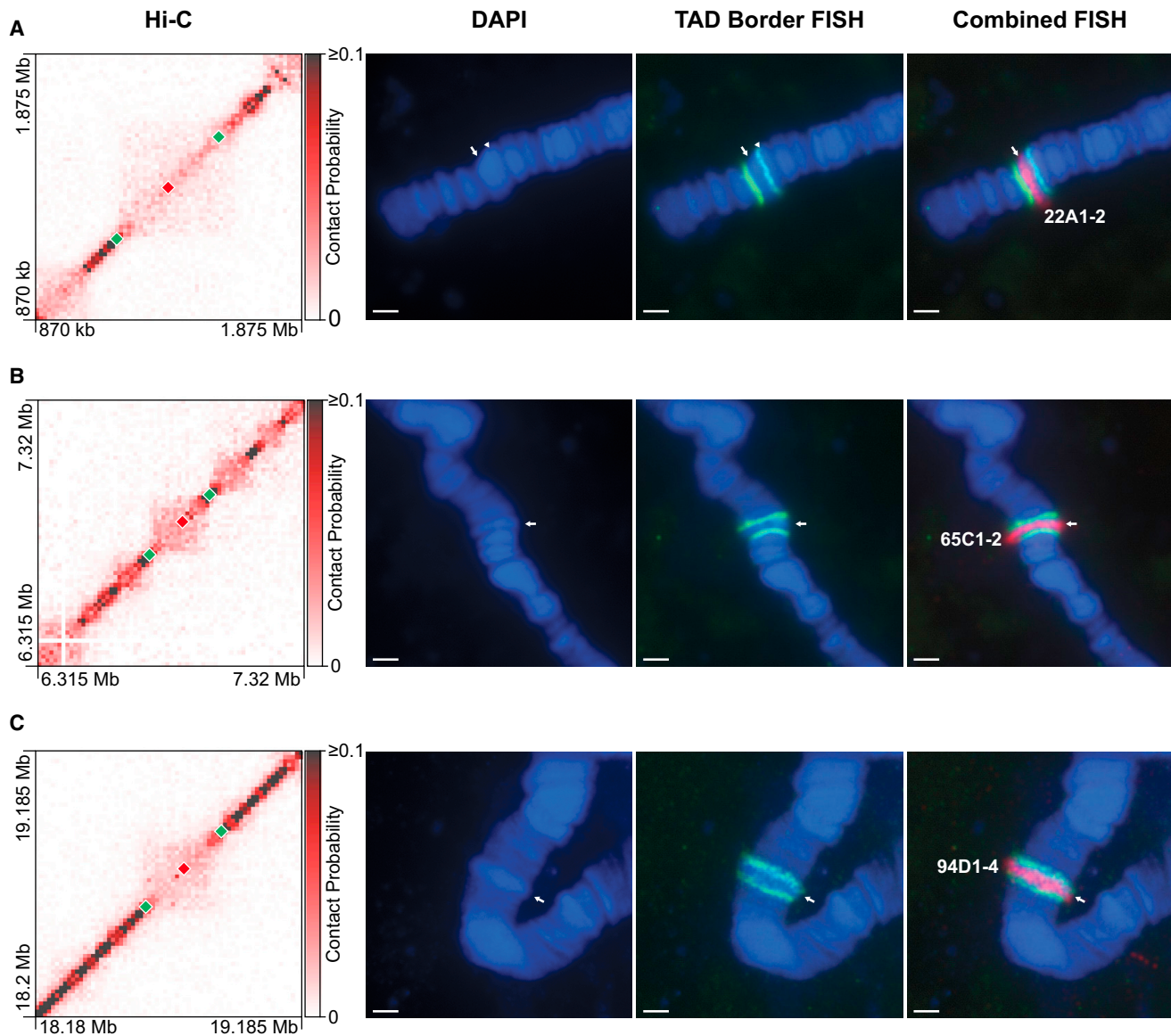
instar larval stage showed no evidence of TADs, but rather strong signals restricted to the diagonal (Figure 3A). Furthermore, quantitation of the interaction pattern of puffs showed a mostly uniform directionality index centered around zero, rather than a TAD-like bias in directionality (Figure 3B). The absence of TADs from puffed bands strengthens the connection between TADs and chromatin condensation.

Hi-C Predicts the Location of Bands and Interbands

As a further test of the relationship between TADs and bands, we asked whether the DNA sequence of a TAD could be used to predict the location of a band for which reliable DNA sequence information was not available. We designed fluorescent probes to the centers and borders of three TADs and hybridized them individually to polytene chromosomes (Figure 4). In every case, the FISH signal from a TAD center probe precisely overlapped with a polytene band. In five of six cases, the FISH signal from a TAD border probe overlapped with the adjacent polytene interband. In the one case where the TAD border probe overlapped with a band (Figure 4A; centromere proximal probe), the overlap occurred at the margin of the band immediately adjacent to a small interband (between bands 22A1-2 and 22A3 in Bridges's map) (Bridges, 1935; Lefevre, 1976). Inasmuch as only 5% of DNA is located within interbands (Beermann, 1972), our success rate of 0.833 in identifying interbands is much greater than ex-

Figure 4. Hi-C Predicts the Location of Polytene Bands

The locations of TADs (normalized Hi-C heatmaps, 15 kb bins; left column) were used to generate FISH probes against TAD centers (red diamonds) or TAD borders (green diamonds), which were hybridized to polytene chromosome spreads counterstained with DAPI (middle-left column). The TAD border FISH signal and the TAD center FISH signal are pseudocolored green and red, respectively, in the merged images (rightmost two columns) and the identity of the polytene band is indicated (right column). White arrows indicate the polytene band of interest.

(A) FISH against a region from chromosome 2L. White arrowhead indicates a small interband between bands 22A1-2 and 22A3 in Bridges's map.

(B) FISH against a region from chromosome 3L.

(C) FISH against a region from chromosome 3R.

Loci presented here are independent of those analyzed in Figure 2. Scale bars, 2 μm.

Lack of Compartments in the Polytene Nucleus

Compartmentation refers to the preferential association of distant chromosomal loci in two groups, which correlate with transcriptional activity and inactivity. At large length scales, chromosomes exhibit polymeric behavior: loci separated by large linear distances are less likely to be in close spatial proximity than loci separated by small distances (Lieberman-Aiden et al., 2009). Associations of distant loci become apparent when a correction is applied for polymeric behavior: the observed number of interactions between two loci is divided by the number of interactions expected due to variations in polymer conformation. The resulting observed/expected heatmap (Figure 6, middle column) displays regions of more or less interaction than the expected chromosome-wide average. The delineation of such regions can be enhanced by correlation analysis, because neighboring loci along the genome in close spatial

A Chromosome X - Polytene Bands Boxed

A Chromosome X - Polytene TADs Boxed

B Chromosome X - Diploid Bands Boxed

B Chromosome X - Diploid TADs Boxed

C

D Diploid TADs   Between Diploid TADs

E
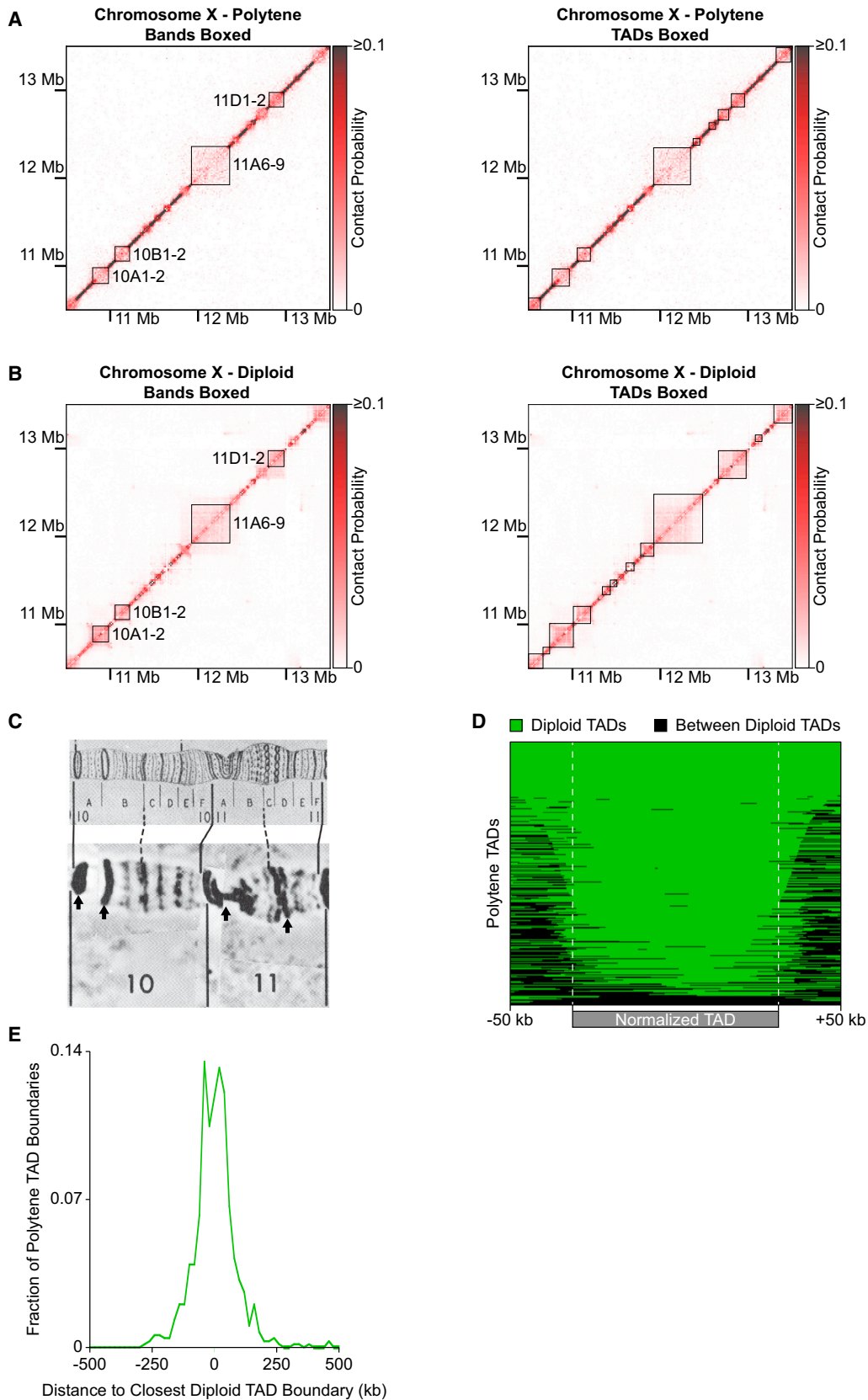
(legend on next page)

proximity should share the same interaction preferences and spatially distant loci should differ in their interaction preferences. A heatmap of the correlations between interaction profiles shows a "plaid" pattern if there are both shared and divergent interaction preferences (Figures 6, right, and S7) (Lieberman-Aiden et al., 2009). Boxes in the plaid pattern are of two types, boxes along the diagonal, at the locations of TADs in the observed heatmap, and off-diagonal boxes, revealing enriched or depleted long-range interactions. The occurrence of off-diagonal boxes is indicative of compartments. Correlation heatmaps computed from published Hi-C data for Kc167 cells and late-stage *Drosophila* embryos (Figures 6, top two rows, and S7) exhibit a plaid pattern and thus compartments, as previously observed (Sexton et al., 2012). By contrast, a correlation heatmap computed from our Hi-C data for *Drosophila* polytene chromosomes shows no evidence of a plaid pattern (Figures 6, bottom row, and S7), indicating an absence of long-range interactions and a lack of compartments defined on that basis. A lack of compartments is not necessarily a consequence of homolog pairing, because homologs are paired in both embryonic (Fung et al., 1998) and Kc (Williams et al., 2007) cells.

DISCUSSION

Our finding of an equivalence between polytene bands and TADs has 2-fold significance. It complements the discovery of TADs by chemical cross-linking with their identification by a structural approach, and it shows that chromatin folding inferred from cross-linking corresponds to bona fide chromosome condensation, to a state so dense it can be seen by light microscopy. Polytene bands are stable from cell to cell (Painter, 1933), as well as when observed in real time (Hochstrasser et al., 1986); TADs are therefore similarly stable. Interbands correspond to regions between TADs, which thus reflect a stable state of chromosome decondensation. The band-interband pattern is similar between polytene nuclei of different tissues (Beermann, 1972). TADs evidently represent a largely invariant, or conserved, feature of chromosome structure. The equivalence of polytene, diploid, and embryonic TADs shows that chromosome structure manifest in the polytene state is general, applicable to the diploid interphase nucleus, and nearly constant among all cell types.

Although TADs are stable between embryonic stem (ES) cells and ES cell-derived lineages, interactions within and between TADs may change during ES cell differentiation (Dixon et al., 2015). So TADs themselves do not directly regulate transcription, but rather additional features, such as transient chromatin loops, are likely presumably involved. The equivalence of polytene bands with TADs indicates that the role of TADs is most likely for the compaction of DNA in the interphase nucleus.

Fine mapping of regions of polytene chromosomes (Zhimulev et al., 2014) has shown that interbands contain regulatory regions and promoters of genes expressed in most tissues, cell lines, developmental stages, and treatment conditions—so-called "housekeeping genes". Regulatory regions, transcription start sites, and 5′ transcribed regions are located in interbands, and the remaining coding portions of genes reside in the adjacent gray bands (Zhimulev et al., 2014). Only half of total RNA synthesis occurs in puffs (Zhimulev and Belyaeva, 1975); the other half may occur in gray bands. Although the moderate resolution of our Hi-C analysis and the lack of comprehensive epigenomic profiling in salivary gland tissues preclude the direct detection of active and inactive TADs in polytene chromosomes, our identification of bands as TADs, along with the characteristics of gray bands, implies a connection between loosely compact gray bands and "active TADs" reported by others (Rao et al., 2014; Sexton et al., 2012). TADs and the regions between them are conserved across cell types likely because transcription is regulated and initiated in interbands and many transcribed genes are active in all cell types.

Conserved interbands or regions between TADs may correspond not only to regulatory regions involved in gene activity but also to those needed for gene repression across cell types. The 5′-regulatory region of *Notch*, which is transcriptionally inactive in salivary glands, lies in an interband between bands 3C6 and 3C7, with the coding region in band 3C7 (Rykowski et al., 1988). The *Notch* mutation *facet-strawberry* (*fa^{swb}*) results in a small deletion harboring an insulator element (Vazquez and Schedl, 2000), which maps to the 3C6–3C7 interband (Rykowski et al., 1988). In the *fa^{swb}* mutation, the 3C6–3C7 interband disappears, resulting in fusion of bands 3C6 and 3C7 (Keppy and Welshons, 1977). Ectopic insertion of the DNA sequence comprising the *fa^{swb}* deletion was necessary and sufficient to split an endogenous band in two and form an interband at the

Figure 5. Conservation of Polytene and Diploid TADs

Normalized Hi-C heatmaps (15 kb bins) of a 3 Mb region of the X chromosome. Left panels: areas bounded by black boxes represent locations of polytene bands for which reliable DNA sequence coordinates are available (Belyaeva et al., 2012; Vatolina et al., 2011). Right panels: areas bounded by black boxes represent TADs.

(A) Hi-C heatmap from polytene cells.

(B) Hi-C heatmap from diploid Kc167 cultured cells.

(C) Photographic image (bottom) of the region of polytene chromosome X from (A) and (B) and the same region of Bridges's chromosome map (top). Arrows indicate bands represented by black boxes in (A) and (B). The DNA sequence coordinates of other bands in this region are not known and therefore cannot be compared with the Hi-C data. Adapted from Lefevre (1976). To prepare the panel, the original photographic image from Lefevre was digitally scanned and then cropped to display the region of the chromosome corresponding to that of the Hi-C analysis in (A) and (B). Black arrows were then superimposed onto the cropped image using Adobe Illustrator to mark the polytene bands represented by black boxes in (A) and (B).

(D) Heatmap of the agreement between polytene TADs and diploid TADs. Each polytene TAD is represented by a row. Green segments overlap with diploid TADs, black segments overlap with regions between diploid TADs. TADs were normalized to the same length and 50 kb of flanking DNA is shown next to each normalized TAD.

(E) Fraction of polytene TAD boundaries (n = 692) at the distance indicated on the abscissa from the closest diploid TAD boundary (calculated in 20 kb windows). See also Figures S4, S5, and S6.
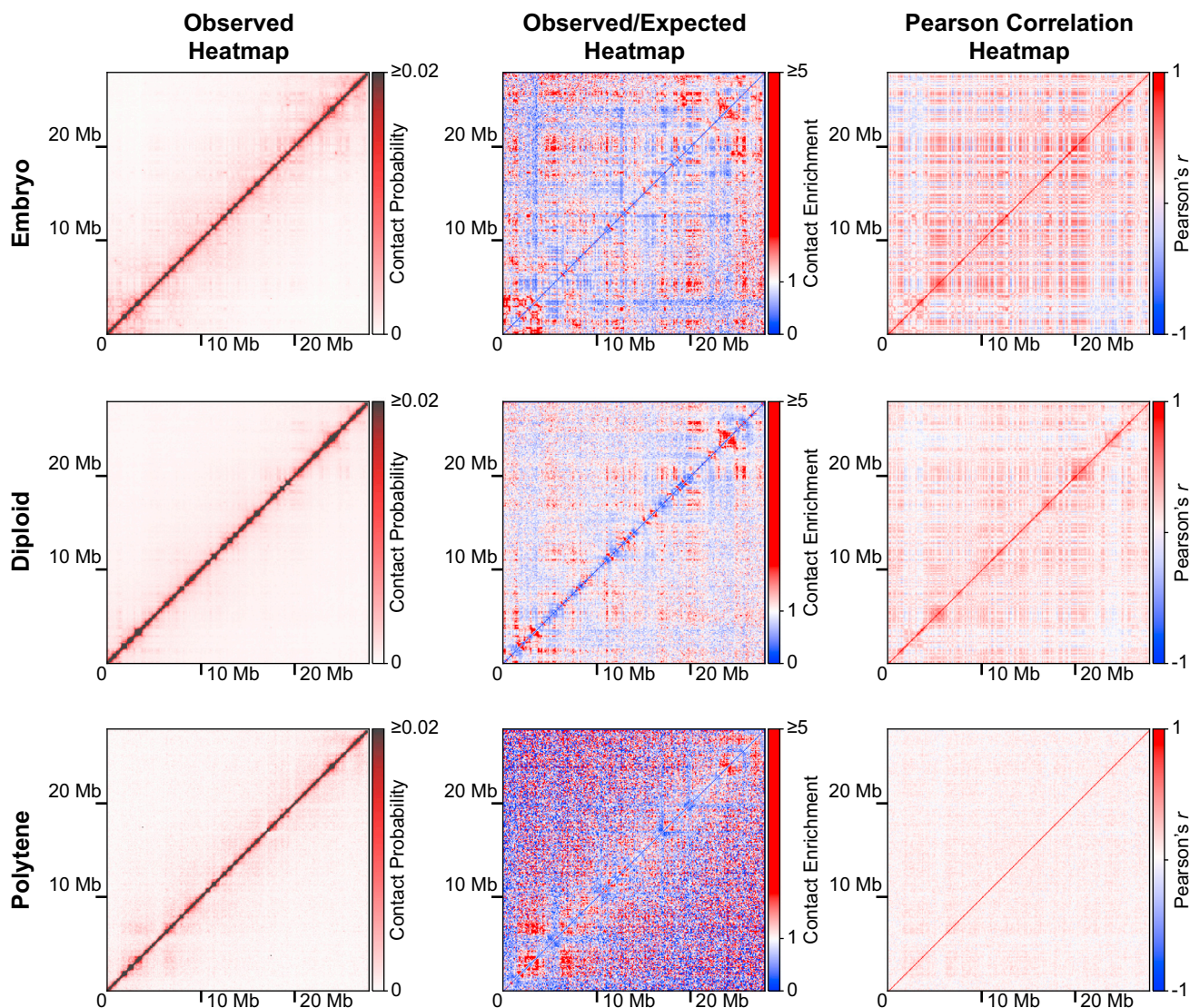
Figure 6. Lack of Compartments in the Polytene Nucleus

Normalized observed (left column), observed/expected (middle column), and Pearson correlation (right column) Hi-C heatmaps (100 kb bins) for embryonic (top row), diploid Kc167 cell (middle row), and polytene (bottom row) chromosome 3R. In the observed/expected heatmaps, interactions less than the expected chromosome-wide average are blue, those greater than the expected chromosome-wide average are red. A plaid pattern in a Pearson correlation heatmap indicates the presence of compartments.

See also Figure S7.

ectopic site (Andreyenkov et al., 2010). Although a factor that binds to the *fa^{swb}* insulator has not yet been identified, maintaining the chromatin in a fully extended state would facilitate protein binding. If the same gene or set of genes is transcriptionally repressed across cell types, the state of chromatin and accessibility to transcriptional repressors may also be conserved, further accounting for the similarity in Hi-C results across cell types.

From the DNA sequences of bands and interbands and measurements on micrographs of polytene chromosomes, the packing ratios (length of DNA to length of chromatin) of bands and interbands may be determined. The ratios range from 158:1 to 205:1 for IH bands, 12:1 to 73:1 for gray bands, and 5:1 to 12:1 in interbands (Rykowski et al., 1988; Vatolina et al., 2011). Conservation of Hi-C results between polytene and diploid cells allows us to deduce the structural states of chromatin, which we refer to as black, gray, and white, respectively, in the interphase, diploid nucleus. The packing ratio of a fully open chain of nucleosomes is expected to be 6.8:1 (Kornberg, 1974), so white chromatin evidently contains fully extended chromatin fibers. Approximately 5% of *Drosophila* DNA, containing promoters and regulatory regions, resides in this fully extended state (Beermann, 1972; Hou et al., 2012). Grey chromatin is up to 10-fold more compact and contains approximately a quarter

Black chromatin
Packing ratio between 150:1 and 200:1
Up to 30-fold more condensed than open chromatin
~70% of total DNA
DNA sequences in inactive TADs

Heterochromatin

Euchromatin

White chromatin
Packing ratio of 7:1
Fully extended chromatin fibers
~5% of total DNA
DNA sequences between TADs

Grey chromatin
Packing ratio between 10:1 and 80:1
Up to 10-fold more condensed than open chromatin
~25% of total DNA
DNA sequences in active TADs

Figure 7. Chromosome Condensation in the Interphase Nucleus

Left: thin section electron micrograph of a nucleus (Cross and Mercer, 1993), with lightly staining euchromatin in the nuclear interior, surrounded by darkly staining heterochromatin, concentrated at the nuclear periphery. Right: cartoon representation of white, gray, and black chromatin, showing proposed relationships to heterochromatin, euchromatin, and the nuclear envelope (yellow). Active TADs in the euchromatin are nearby other active TADs and inactive TADs in the heterochromatin are nearby other inactive TADs, resulting in gray-gray and black-black TAD-TAD interactions. The actual pattern of chromatin folding is unknown and indicated only schematically.

of the *Drosophila* genome (Filion et al., 2010; Sexton et al., 2012), within which are located the coding regions of active genes. Therefore, transcriptional regulation occurs in a fully extended state and elongation occurs in a partially compacted state. Black chromatin is up to 30-fold more condensed than fully extended chromatin fibers. It contains inactive genes and accounts for the remaining 70% of the genome.

The distribution of interactions within a TAD is informative about the pattern of chromatin condensation. The distribution is remarkably uniform, with every site in a TAD, particularly inactive TADs (Sexton et al., 2012), almost equally likely to become cross-linked with every other site (Figure 6B). Additionally, since homologs are paired in nearly all cell types (Fung et al., 1998; Williams et al., 2007), *Drosophila* chromosomes are particularly illuminating with regard to chromatin condensation. Although cytology clearly demonstrates homologous chromosome pairing, our Hi-C analysis indicates that, even for polytene chromosomes, at the molecular level they are not perfectly aligned. Nucleosome by nucleosome, the chromatin fibers from homologs and chromatids are not in lockstep—there is some variability in the path of each of the chromatin fibers. Although at the cytological level, bands are present at reproducible positions, there is some variation at the molecular level in the path of the DNA. Taken together, the interactions revealed by Hi-C cannot arise primarily from specific, stable chromatin loops, but rather the pattern of condensation within gray and black chromatin must vary from one cell to another within a population,

consistent with conclusions of others (Gibcus and Dekker, 2013). There can be no unique pattern of condensation.

Studies of chromosomal protein distribution and histone modifications have revealed up to 16 functional classes of chromatin (Ernst et al., 2011; Filion et al., 2010; Ho et al., 2014; Kharchenko et al., 2011; Ram et al., 2011; Rao et al., 2014; Sexton et al., 2012), and these classes are conserved between polytene and diploid cells (Zhimulev et al., 2014). One class corresponds closely with interbands, others classes with the flanking bands (Zhimulev et al., 2014). Structural information from polytene chromosomes identifies three structural states – black, gray, and white. A structural state may encompass multiple functional classes. For example, at the current level of resolution, we cannot distinguish between polycomb and HP1 repressed heterochromatin.

The equivalence of polytene bands with polytene TADs and the virtual identity of polytene TADs with diploid TADs imply a close correspondence of bands with diploid TADs. The structure of interphase chromosomes revealed by light microscopy of polytene nuclei therefore applies to the organization of the diploid cell nucleus. A subset of TADs in mammalian cells associates with the nuclear lamina (Dixon et al., 2012), and condensed chromatin is mostly located at the periphery of nuclei in electron micrographs of virtually all fixed, embedded, heavy metal-stained eukaryotic cell preparations (Figure 7, left). Polytene bands are also identified by heavy metal staining in electron micrographs and interactions of polytene chromosomes with the nuclear envelope, although infrequent, are almost entirely confined to the IH bands (Hochstrasser et al., 1986). If inactive TADs (black chromatin) are often located near the nuclear periphery, then the regions between inactive TADs, consisting of active TADs and fully extended chromatin fibers (white and gray chromatin), must loop into the interior of the nucleus (Figure 7, right). It follows that inactive TADs (black chromatin) correspond to classical heterochromatin (condensed chromatin at the nuclear periphery) and active TADs and the regions between

TADs (white and gray chromatin) correspond to classical euchromatin (less dense chromatin in the nuclear interior). On the basis of the measurements mentioned above, heterochromatin at the nuclear periphery consists of up to 30-fold condensed chromatin fibers, while euchromatin in the nuclear interior contains both 10-fold condensed and fully extended fibers. The DNA sequences that reside in each of these regions are known from the results of Hi-C analysis. The identification of polytene TADs thus gives meaning to the familiar picture of the interphase nucleus.

Compartments can be understood in terms of the physical picture of the interphase nucleus. TADs and compartments are related, but differ in origin and significance. TADs are aligned with the off-diagonal boxes in the plaid pattern that defines compartments; the boundaries of TADs define the rows and columns of the plaid pattern (Figure 6, right column). Therefore the off-diagonal boxes arise from interactions between TADs. The difference is that TADs are due to persistent interactions. They reflect a stable state of axial condensation of the chromatin fiber. By contrast, off-diagonal boxes are most apparent following correction for random polymer conformation and sharpening of the corrected heatmap. Off-diagonal boxes are therefore due to transient contacts between distant regions, for example short-lived contacts between TADs in neighboring euchromatic chromatin in the nuclear interior (Figure 7, right). This picture explains why off-diagonal boxes fall in two categories of interacting sequences, corresponding to active and inactive genes: active TADs (gray chromatin) tend to be in the vicinity of other active TADs and inactive TADs (black chromatin) in the vicinity of other inactive TADs. In observed heatmaps, the off-diagonal boxes usually appear uniform in signal strength over their entire area, indicative of roughly equal likelihood of interaction of every sequence in one TAD with every sequence in another (Sexton et al., 2012). The reason is probably because the TADs themselves are condensed in a variable manner, as discussed above. All sequences have roughly equal probabilities of appearing on the surface of the condensed structure and interacting with adjacent condensed regions. Moreover, a single TAD may make many distant contacts, so there cannot be a unique pattern of TAD-TAD interactions, but rather the trajectory of a chromosome must vary from one cell to another.

Compartments refer to loci grouped on the basis of frequency of interaction. Compartments are not regions with boundaries, in the conventional sense of the term. They correspond to areas of the nucleus, for instance the interior and the periphery, but the areas themselves do not determine the state of gene activity (Therizols et al., 2014). The properties of polytene chromosomes are illustrative. Polytene chromosomes possess both active and inactive genes, but they exhibit no compartments, and they reside mostly in the nuclear interior, making little or no contact with the periphery. They lack compartments because they do not fold upon themselves to a significant extent and therefore have no TAD-TAD interactions. TADs may occur without compartments, and chromosome condensation and gene regulation do not require compartments. Condensation along the chromosome axis is conserved, as shown by the equivalence of polytene, diploid, and embryonic TADs, whereas compartments may vary, depending on cell type and state of transcriptional activity.

The organization of the interphase nucleus in *Drosophila* is relevant to the mouse and to humans, where TADs organize chromosomes into spatial modules connected by short chromatin segments (Dixon et al., 2012). Furthermore, biochemical fractionation of open chromatin fibers from human cells revealed that the fibers are cytologically decondensed (Gilbert et al., 2004), and it is now apparent that these fibers are likely in the fully extended state. The packing ratios, DNA sequences, functional states, and chromosomal protein patterns of the differentially staining areas of the interphase nucleus are thus determined. Genome-wide amplification and alignment in the polytene state reveals interphase chromosome structure at the level of light microscopy, likely applicable to the diploid state in all monocentric metazoans.

EXPERIMENTAL PROCEDURES

Hi-C
Hi-C was performed using a tethering approach (Kalhor et al., 2012) to improve the signal-to-noise from a limited amount of manually dissected, primary tissue. In brief, *Drosophila melanogaster* third instar larvae salivary glands were manually dissected and fixed with 2% EM grade paraformaldehyde. Cross-linked proteins were then biotinylated at cysteine residues and the DNA digested with DpnII. Digested chromatin was bound to streptavidin beads, thoroughly washed to remove uncross-linked DNA, DNA ends filled in with biotin-14-dATP, and free DNA ends ligated together. DNA-protein cross-links were reversed, DNA purified, biotinylated nucleotides marking unligated ends removed, and then DNA sheared to a mean size of ∼200 bp. The biotinylated DNA was pulled down with streptavidin beads, prepared for and subjected to high-throughput Illumina sequencing. Further details provided in the Supplemental Experimental Procedures.

Hi-C Analysis
Hi-C reads were mapped to the dm3 reference genome using Bowtie 2 and assigned to DpnII restriction fragments. Reads mapping to the same restriction fragment, separated by less than the library insert size, within 4 bp of a restriction site, and duplicate reads were removed. Exceptionally large (>100 kb) and small (<100 bp) restriction fragments and fragments with the highest 0.5% of counts were also removed. Filtered fragments were assigned to 15 kb genomic bins, unless otherwise indicated. Further filtering at the bin level removed bins where less than half the bin was sequenced, the lowest 1% of bins, and the highest 0.05% of interchromosomal bins. The resulting Hi-C heatmaps were normalized using a previously described iterative approach (Imakaev et al., 2012). All of the above steps were performed using a previously described pipeline (Imakaev et al., 2012).

The genome-wide directionality index (DI), a modified chi-square statistic to measure the directional interaction bias of a locus, was determined as previously described (Dixon et al., 2012). TADs were identified by using the value midway between the mean of the values in the lowest two deciles of the contact probability along the diagonal and the mean of the values in the highest two deciles of the contact probability along the diagonal as a threshold for a low-pass filter. TADs were further required to have a minimum size of 75 kb to generously satisfy the Nyquist sampling criterion. Statistical comparisons between polytene bands and TADs or between TADs from different cell-types employed a bootstrapping approach to determine the significance of (1) the overlap between the features in the two lists, (2) the aggregate unmatched length of the features in the two lists, and (3) the Euclidean distance between the corners of the nearest features in the two lists.

The presence or absence of Hi-C compartments was determined for each chromosome as previously described (Lieberman-Aiden et al., 2009) by dividing the observed heatmap by the expected heatmap empirically determined by dividing the number of observed interactions at a given distance

by the total number of loci separated by that same distance. The Pearson correlation coefficient between the ith row and j^{th} column of the observed/expected heatmap gives the Pearson correlation heatmap.

Hi-C data and TADs from Kc167 cells (Hou et al., 2012) and embryonic nuclei (Sexton et al., 2012) were previously published and obtained from GEO: GSE38468 and GEO: GSE34453, respectively. Further details provided in the Supplemental Experimental Procedures.

FISH

FISH was performed on acid-fixed, squashed salivary glands as previously described (Kennison, 2000; Pardue, 2000) with further details provided in the Supplemental Experimental Procedures. Primers used to generate FISH probes are listed in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The Hi-C data reported in this study are deposited in GEO: GSE72512.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.026.

AUTHOR CONTRIBUTIONS

K.P.E. and R.D.K. designed the study. K.P.E. performed Hi-C experiments and data analysis. T.A.H. prepared tissue for Hi-C and provided *Drosophila* expertise. K.P.E. performed FISH with guidance from T.A.H. All authors discussed the results. K.P.E. and R.D.K. wrote the manuscript with input from T.A.H.

ACKNOWLEDGMENTS

REFERENCES

Agard, D.A., and Sedat, J.W. (1983). Three-dimensional architecture of a polytene nucleus. Nature *302*, 676–681.

Andreyenkov, O.V., Volkova, E.I., Demakov, S.A., Semeshin, V.F., and Zhimulev, I.F. (2010). The decompact state of interchromomeric chromatin from the 3C6/C7 region of Drosophila melanogaster is determined by short DNA sequence. Dokl. Biochem. Biophys. *431*, 57–59.

Ashburner, M. (1967). Patterns of puffing activity in the salivary gland chromosomes of Drosophila. I. Autosomal puffing patterns in a laboratory stock of Drosophila melanogaster. Chromosoma *21*, 398–428.

Balbiani, E.G. (1881). Sur la structure du noyau des cellules salivaires chez les larves de Chironomus. Zool. Anz. *iv*, 637–641.

Beermann, W. (1972). Chromomeres and genes. Results Probl. Cell Differ. *4*, 1–33.

Belyaeva, E.S., Goncharov, F.P., Demakova, O.V., Kolesnikova, T.D., Boldyreva, L.V., Semeshin, V.F., and Zhimulev, I.F. (2012). Late replication domains in polytene and non-polytene cells of Drosophila melanogaster. PLoS ONE *7*, e30035.

Bridges, C.B. (1935). Salivary chromosome maps with a key to the banding of the chromosomes of Drosophila melanogaster. J. Hered. *26*, 60–64.

Cross, P.C., and Mercer, K.L. (1993). Cell and Tissue Ultrastructure: A Functional Perspective (W.H. Freeman).

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., and van Steensel, B. (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212–224.

Flemming, W. (1882). Zellsubstanz, Kern und Zelltheilung (Leipzig: F. C. W. Vogel).

Fung, J.C., Marshall, W.F., Dernburg, A., Agard, D.A., and Sedat, J.W. (1998). Homologous chromosome pairing in Drosophila melanogaster proceeds through multiple independent initiations. J. Cell Biol. *141*, 5–20.

Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. Mol. Cell *49*, 773–782.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. (2004). Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. Cell *118*, 555–566.

Ho, J.W.K., Jung, Y.L., Liu, T., Alver, B.H., Lee, S., Ikegami, K., Sohn, K.-A., Minoda, A., Tolstorukov, M.Y., Appert, A., et al. (2014). Comparative analysis of metazoan chromatin organization. Nature *512*, 449–452.

Hochstrasser, M., Mathog, D., Gruenbaum, Y., Saumweber, H., and Sedat, J.W. (1986). Spatial organization of chromosomes in the salivary gland nuclei of Drosophila melanogaster. J. Cell Biol. *102*, 112–123.

Hou, C., Li, L., Qin, Z.S., and Corces, V.G. (2012). Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Mol. Cell *48*, 471–484.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods *9*, 999–1003.

Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat. Biotechnol. *30*, 90–98.

Kennison, J.A. (2000). Preparation and analysis of polytene chromosomes. In Drosophila Protocols, W. Sullivan, M. Ashburner, and R.S. Hawley, eds. (Cold Spring Harbor Laboratory Press), pp. 111–117.

Keppy, D.O., and Welshons, W.J. (1977). The cytogenetics of a recessive visible mutant associated with a deficiency adjacent to the notch locus in Drosophila melanogaster. Genetics *85*, 497–506.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., et al.

(2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature *471*, 480–485.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868–871.

Lefevre, G., Jr. (1976). A photographic representation and interpretation of the polytene chromosomes of Drosophila melanogaster salivary glands. In The Genetics and Biology of Drosophila, M. Ashburner and E. Novitski, eds. (Academic Press), pp. 31–66.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251–260.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Painter, T.S. (1933). A new method for the study of chromosome aberrations and the plotting of chromosome maps. Science *78*, 585–586.

Pardue, M.-L. (2000). In situ hybridization to polytene chromosomes. In Drosophila Protocols, W. Sullivan, M. Ashburner, and R.S. Hawley, eds. (Cold Spring Harbor Laboratory Press), pp. 119–129.

Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M., et al. (2011). Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. Cell *147*, 1628–1639.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rykowski, M.C., Parmelee, S.J., Agard, D.A., and Sedat, J.W. (1988). Precise determination of the molecular limits of a polytene chromosome band: regulatory sequences for the Notch gene are in the interband. Cell *54*, 461–472.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Therizols, P., Illingworth, R.S., Courilleau, C., Boyle, S., Wood, A.J., and Bickmore, W.A. (2014). Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. Science *346*, 1238–1242.

Urata, Y., Parmelee, S.J., Agard, D.A., and Sedat, J.W. (1995). A three-dimensional structural dissection of Drosophila polytene chromosomes. J. Cell Biol. *131*, 279–295.

Vatolina, T.Y., Boldyreva, L.V., Demakova, O.V., Demakov, S.A., Kokoza, E.B., Semeshin, V.F., Babenko, V.N., Goncharov, F.P., Belyaeva, E.S., and Zhimulev, I.F. (2011). Identical functional organization of nonpolytene and polytene chromosomes in Drosophila melanogaster. PLoS ONE *6*, e25960.

Vazquez, J., and Schedl, P. (2000). Deletion of an insulator element by the mutation facet-strawberry in Drosophila melanogaster. Genetics *155*, 1297–1311.

Williams, B.R., Bateman, J.R., Novikov, N.D., and Wu, C.-T. (2007). Disruption of topoisomerase II perturbs pairing in Drosophila cell culture. Genetics *177*, 31–46.

Zhimulev, I.F., and Belyaeva, E.S. (1975). 3H-Uridine labelling patterns in the Drosophila melanogaster salivary gland chromosomes X, 2R and 3L. Chromosoma *49*, 219–231.

Zhimulev, I.F., Semeshin, V.F., Kulichkov, V.A., and Belyaeva, E.S. (1982). Intercalary heterochromatin in Drosophila. I. Localization and general characteristics. Chromosoma *87*, 197–228.

Zhimulev, I.F., Zykova, T.Y., Goncharov, F.P., Khoroshko, V.A., Demakova, O.V., Semeshin, V.F., Pokholkova, G.V., Boldyreva, L.V., Demidova, D.S., Babenko, V.N., et al. (2014). Genetic organization of interphase chromosome bands and interbands in Drosophila melanogaster. PLoS ONE *9*, e101631.

Cell

Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes

Graphical Abstract

Highlights

- Linear tracking within CBE-based loops restricts RAG off-target activity genome-wide

- ATM deficiency allows RAG-initiated DSBs to escape from loop domains and translocate

- RAG off-target activity marks certain *IgH* locus V(D)J recombination domains

- Robust RAG off-targets occur in pairs in highly accessible regions within a loop
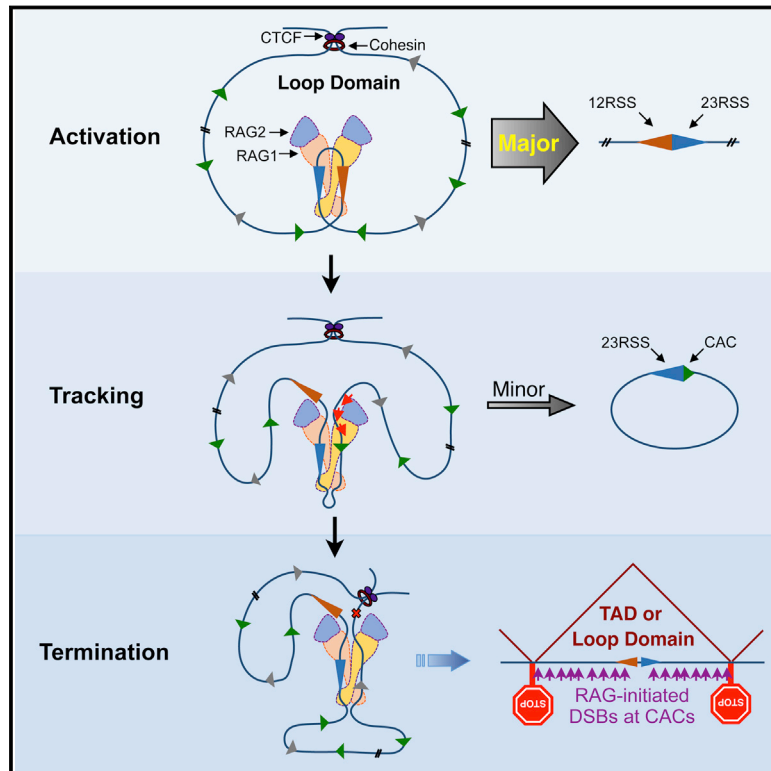
Authors

Jiazhi Hu, Yu Zhang, Lijuan Zhao, ..., Fei-long Meng, David G. Schatz, Frederick W. Alt

Correspondence

alt@enders.tch.harvard.edu

In Brief

The 3D architecture of the genome harnesses RAG endonuclease activity by allowing it to directionally locate convergent recombination signal sequences and related CAC-containing motifs within individual chromatin loop domains.

Accession Numbers

GSE73007

CellPress

Article

Cell

Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes

Jiazhi Hu,[1,2,3,5] Yu Zhang,[1,2,3,5] Lijuan Zhao,[1,2,3] Richard L. Frock,[1,2,3] Zhou Du,[1,2,3] Robin M. Meyers,[1,2,3] Fei-long Meng,[1,2,3] David G. Schatz,[1,4] and Frederick W. Alt[1,2,3,*]

[1]Howard Hughes Medical Institute
[2]Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA 02115, USA
[3]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[4]Department of Immunobiology, Yale University School of Medicine, 300 Cedar Street, Box 208011, New Haven, CT 06520-8011, USA
[5]Co-first author
*Correspondence: alt@enders.tch.harvard.edu
http://dx.doi.org/10.1016/j.cell.2015.10.016

SUMMARY

RAG initiates antibody V(D)J recombination in developing lymphocytes by generating "on-target" DNA breaks at matched pairs of bona fide recombination signal sequences (RSSs). We employ bait RAG-generated breaks in endogenous or ectopically inserted RSS pairs to identify huge numbers of RAG "off-target" breaks. Such breaks occur at the simple CAC motif that defines the RSS cleavage site and are largely confined within convergent CTCF-binding element (CBE)-flanked loop domains containing bait RSS pairs. Marked orientation dependence of RAG off-target activity within loops spanning up to 2 megabases implies involvement of linear tracking. In this regard, major RAG off-targets in chromosomal translocations occur as convergent RSS pairs at enhancers within a loop. Finally, deletion of a CBE-based *IgH* locus element disrupts V(D)J recombination domains and, correspondingly, alters RAG on- and off-target distributions within *IgH*. Our findings reveal how RAG activity is developmentally focused and implicate mechanisms by which chromatin domains harness biological processes within them.

INTRODUCTION

During B and T lymphocyte development, exons encoding antigen-binding immunoglobulin (Ig) or T cell receptor (TCR) variable regions are assembled from V, D, and J gene segments by V(D)J recombination (Alt et al., 2013). V(D)J recombination is initiated by RAG endonuclease, which introduces DNA double-stranded breaks (DSBs) between a pair of V, D, and J coding gene segments and their flanking recombination signal sequences (RSSs) (Schatz and Swanson, 2011). A bona fide RSS comprises a conserved palindromic heptamer represented by the canonical CACAGTG sequence, a degenerate spacer of 12 or 23 base pairs (bp), and a less-conserved A-rich nonamer (Figure 1A;

Schatz and Swanson, 2011). RSSs with 12- or 23-bp spacers are termed 12RSSs and 23RSSs, respectively. Efficient RAG cleavage is restricted to a pair of participating coding segments flanked, respectively, by a 12RSS and a 23RSS, referred to here as paired bona fide RSSs. This 12/23 RSS restriction helps direct RAG to appropriate targets within antigen receptor loci (Alt et al., 2013).

RAG cleavage generates a pair of blunt broken RSS ends and a pair of hairpin-sealed coding ends (Figure 1B; Schatz and Swanson, 2011). Classical non-homologous end joining (C-NHEJ) fuses the two RSS ends precisely to form RSS joins and opens the two coding-end hairpins and joins them to form coding joins, which may be "processed" to lose or gain several nucleotides from each end (Figure 1B; Alt et al., 2013). While potential bona fide RSS-related sequences occur frequently across the genome, only a small number are documented RAG off-targets ("cryptic RSSs") (Merelli et al., 2010). Such RAG off-target activity contributes to oncogenic deletions or translocations in immature B and T cell cancers (Boehm et al., 1989; Larmonie et al., 2013; Onozawa and Aplan, 2012; Papaemmanuil et al., 2014). While RAG1 and RAG2 bind several thousand genomic sites that mostly correspond to active promoters and enhancers (Ji et al., 2010; Teng et al., 2015), lower densities of cryptic RSS heptamers near transcription start sites may help to limit RAG off-target activity (Teng et al., 2015).

The mouse *IgH* locus spans 2.7 megabases (Mb) with V_Hs and their downstream 23RSSs embedded in a 2.4-Mb distal portion separated by a 100-kb intergenic region from D_Hs flanked on both sides by 12RSSs and J_Hs flanked upstream by 23RSSs. Even though 12/23 restriction should allow V_Hs to join to un-rearranged D_Hs, *IgH* V(D)J recombination is "ordered," with D_H to J_H joining occurring in early progenitor (pro)-B cells followed by appendage of a V_H to a DJ_H complex (Alt et al., 2013). Ordered rearrangement and other levels of *IgH* V(D)J recombination regulation are mediated by modulating gene segment accessibility to RAG (Yancopoulos and Alt, 1986). In this regard, *IgH* contains a critical regulatory element termed *intergenic control region 1* (IGCR1) within the V_H-to-D_H interval (Guo et al., 2011). IGCR1 suppresses proximal V_H transcription and rearrangement at the D_H-to-J_H joining stage and, thereby, mediates broad levels of V(D)J recombination control, including diversification of antibody repertoires, by indirectly promoting increased utilization of distal
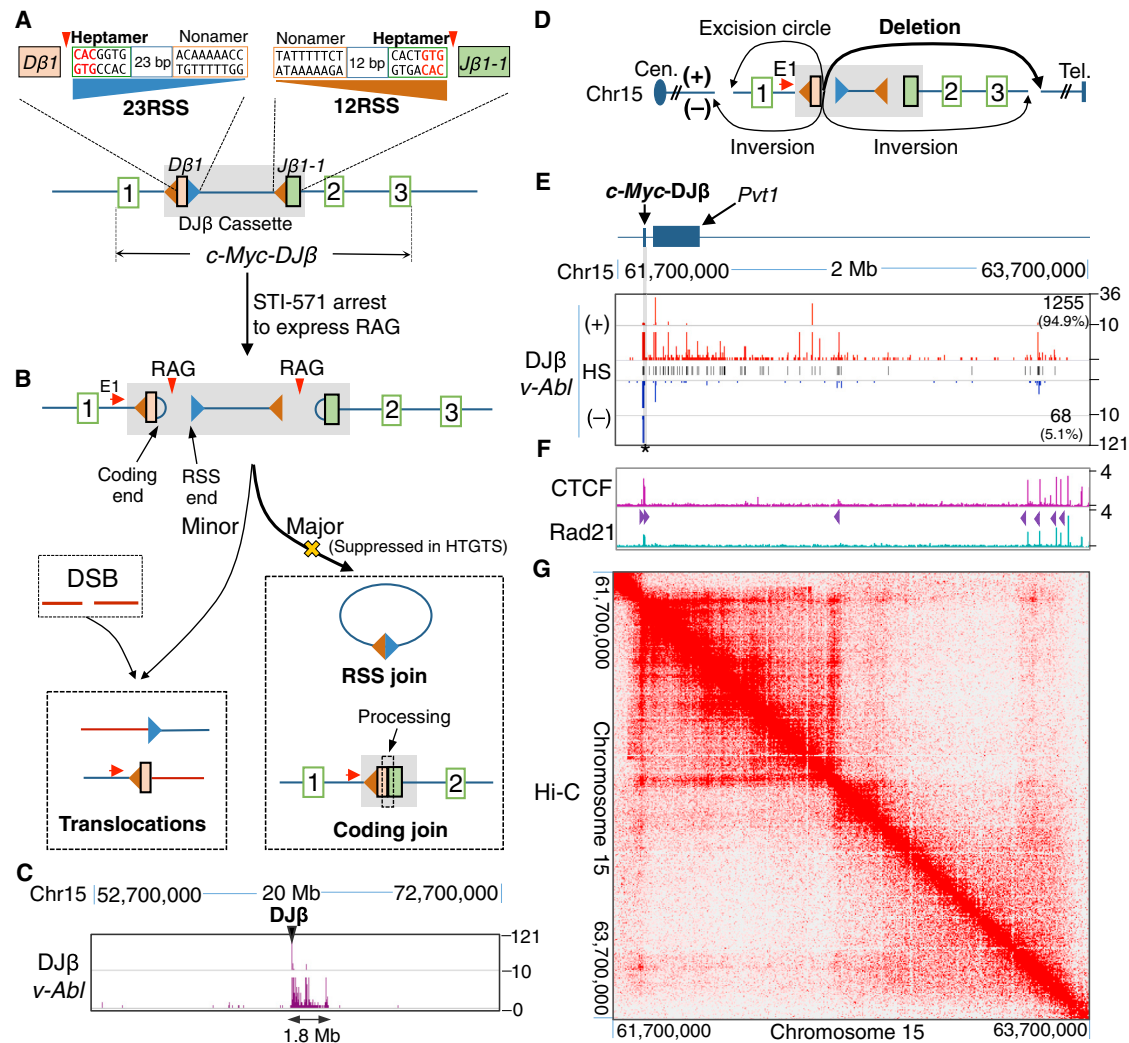
Cell

Figure 1. Abundant DSBs in the 1.8-Mb *c-Myc*-DJβ Loop Domain in *v-Abl* Pro-B Cells

(A) Diagram of *c-My*c-DJβ and sequences of Dβ1 23RSS (blue) and Jβ1-1 12RSS (orange). Red triangles indicate RAG cleavage-sites.

(B) RAG-initiated DSBs in *c-Myc*-DJβ cassette participate in cassette DJβ rearrangements but rarely to translocations involving DSBs on other chromosomes. Red arrows indicate HTGTS primer positions.

(C) Linear plot with a broken y axis showing HTGTS junction profiles in indicated 20-Mb region containing *c-Myc*-DJβ.

(D) Potential junctional outcomes between bait Dβ1 23RSS coding ends and other DSBs in *cis* include deletions, excision circles, and inversions.

(E) HTGTS junction profiles in *v-Abl* cells within indicated 2-Mb region containing *c-Myc*-DJβ. For all panels, unmarked ticks represent 0. Black lines in the middle show hotspot (HS) positions (listed in Table S1). Junction numbers and percentages in + or − orientation downstream of *c-Myc*-DJβ are shown. Cassette location is shadowed in gray. Star indicates a good cryptic RSS.

(F) ChIP-seq profiles of CTCF and Rad21 in the 2-Mb region defined in (E). CBE orientation is indicated by purple triangles.

(G) Heatmap showing the 1.8-Mb *c-Myc* loop domain defined by in situ Hi-C data in CH12-LX cell line.

See also Figure S1 and Table S1.

V_Hs. The most D-proximal V_H (V_H81x), while preferentially utilized in wild-type (WT) pro-B cells (Yancopoulos et al., 1984), is even more frequently utilized upon IGCR1 inactivation (Guo et al., 2011).

The CTCF factor binds directionally to an ∼14-bp DNA target (Nakahashi et al., 2013), referred to as a CTCF-binding element (CBE) (Guo et al., 2011). CTCF is implicated in transcriptional insulation through ability to mediate chromatin loops (Ong and Corces, 2014). IGCR1 function relies on two divergently oriented

CBEs within it (Guo et al., 2011). Besides IGCR1 CBEs, the 3′ *IgH* boundary harbors a CBE cluster (termed "3′ CBEs"), and single CBEs occur just downstream of proximal V_Hs and in intergenic regions between distal V_Hs (Degner et al., 2009). V_H CBEs are convergently oriented with respect to the upstream IGCR1 CBE, and 3′CBEs are convergently oriented with respect to the downstream IGCR1 CBE (Guo et al., 2011). Mutational studies of individual IGCR1 CBEs indicated that loop(s) mediated by the downstream CBE focus RAG activity in early pro-B cells

within a domain containing the D_Hs and J_Hs, while a second domain mediated by the upstream CBE sequesters proximal V_Hs from RAG activity (Lin et al., 2015).

Eukaryotic genomes are organized into a hierarchy of architectures. Hi-C shows that chromatin is organized into topologically associated domains (TADs) that occur on Mb or sub-Mb scales and that have high-frequency chromatin interactions within them (Dixon et al., 2012; Nora et al., 2012). Boundaries of such domains are often co-anchored by long-range interactions of sites bound by CTCF in association with cohesin (Phillips-Cremins et al., 2013; Zuin et al., 2014). Recent higher-resolution in situ Hi-C further revealed that mammalian genomes are divided into contact domains at an average scale of 185 kb (Rao et al., 2014). Contact domains with endpoints that generate a loop are termed loop domains (Rao et al., 2014). Loop domains genome-wide are commonly associated with pairs of convergent CBEs bound by CTCF and cohesin (Rao et al., 2014; Vietri Rudan et al., 2015). TADs have been implicated in replication timing (Pope et al., 2014), super-enhancer-driven transcription (Dowen et al., 2014), and DSB synapsis during antibody class-switch recombination (CSR) (Dong et al., 2015; Zarrin et al., 2007), as well as in promoting normal limb development (Lupiáñez et al., 2015). Mechanistic aspects of how loop domains and TADs function are not well understood.

Our recent studies suggested an unanticipated source of RAG off-target activity within long chromatin domains. To study oncogenic consequences, we generated mice with $Tcr\beta$ Dβ1 and Jβ1-1 segments inserted into intron one of the c-Myc oncogene ("c-Myc-DJβ cassette"). Despite frequent c-Myc-DJβ cassette recombination in developing lymphocytes, these mice do not develop lymphoma (Ranganath et al., 2008). However, ATM-deficient, c-Myc-DJβ cassette mice develop B cell lymphomas with c-Myc translocations/amplifications that fuse RAG-generated IgH DSBs to sequences over a several-hundred-kb region downstream of c-Myc (Tepsuporn et al., 2014). These downstream translocations occur exclusively on the cassette allele but do not involve the cassette, suggesting that RAG activity at bona fide RSS pairs within c-Myc promotes cutting at linked downstream cryptic RSSs (Tepsuporn et al., 2014). On this basis, we identify an immense number of previously unsuspected RAG off-targets generated by a mechanism that has broader implications for gene regulation within loop domains.

RESULTS

HTGTS Assay for RAG On-Target and Off-Target DSBs and Translocations

To test the hypothesis that the c-Myc-DJβ cassette promotes cutting at cryptic RSSs downstream of c-Myc, we generated a v-Abl-transformed pro-B cell line from mice homozygous for the c-Myc-DJβ cassette allele (referred to as "c-Myc-DJβ pro-B line"). In such lines, RAG expression can be induced in the context of G1 cell-cycle arrest following treatment with the v-Abl kinase inhibitor STI-571 (Bredemeyer et al., 2006). Due to propensity of cycling v-Abl transformants to form Dβ1-to-Jβ1-1 cassette rearrangements at low level, we were able to isolate just one v-Abl pro-B clone with an un-rearranged cassette allele (Figure 1A). This clone had a second cassette allele in DJβ-rear-

ranged configuration, which is inert for rearrangement (see below). Upon G1 arrest and RAG expression, the c-Myc-DJβ construct undergoes high-frequency bona fide Dβ1-to-Jβ1-1 rearrangements, which fuse the downstream coding end of Dβ1 (23RSS-associated) to the Jβ1-1 12RSS-associated coding end in the chromosome and, correspondingly fuse the Dβ1 23RSS to the Jβ1-1 12RSS within an excision circle (Figures 1A and 1B). To detect potential cryptic RSSs activated by the c-Myc-DJβ cassette in these v-Abl pro-B cells, we employed high-throughput genome-wide translocation sequencing (HTGTS). HTGTS is a highly sensitive DSB and translocation assay that identifies junctions between a broken end of a fixed "bait" DSB and ends of other prey DSBs genome-wide (Chiarle et al., 2011; Dong et al., 2015; Frock et al., 2015). For these analyses, we used an HTGTS bait primer termed "c-Myc E1" that anneals with sequences 213 bp upstream of the cassette Dβ1 23RSS. This primer detects Dβ1 downstream coding end joins to Jβ1-1 coding ends and to other DSBs genome-wide (Figure 1B).

In the c-Myc-DJβ pro-B line, the vast majority of recovered HTGTS junctions represented expected bona fide cassette Dβ1-to-Jβ1-1 coding joins. To enhance off-target detection, we experimentally suppressed recovery of bona fide cassette DJβ joins (Figure 1B; Supplemental Information). The vast majority of remaining Dβ1 downstream coding-end junctions, representing 1%–3% of total junctions, occurred to sequences up to 1.8 Mb downstream of c-Myc, with additional joins to sequences about 1 kb upstream. Notably, the junctions in this 1.8-Mb region abruptly ended in both directions (Figure 1C; see below). Indeed, the only other clear-cut hotspot region genome-wide occurred at about 0.02% of total junctions and involved low-level translocations to $Ig\kappa$ (Figure S1A), a major bona fide RAG target in v-Abl pro-B cells (Zhang et al., 2012). Approximately 20% of the apparent RAG off-target sites in the 1.8-Mb domain represented recurrent ("hotspot") junctions that, in some cases, were recovered dozens of times in independent libraries (Table S1). HTGTS analysis of bone marrow (BM) pro-B cells from c-Myc-DJβ mice gave similar results (Figures S1A–S1C).

We also isolated an ATM-deficient v-Abl c-Myc-DJβ pro-B line in which one allele had an inversion that joined the Dβ1 23RSS to a cryptic RSS (5′-CACAGTT) in the Jβ1-1 segment (Figure S1D). In this line, the second c-$Myc^{DJ\beta}$ allele was in the inert DJβ configuration. Following G1 arrest, HTGTS employing the c-Myc E1 primer revealed that the major "bona fide" V(D)J joining event in this line (>97% of recovered junctions) was inversional joining of the Dβ1 12RSS (the upstream Dβ1 RSS) to the inverted Dβ1 23RSS 693bp downstream (Figure S1D). The vast majority of remaining joins (~3% of total junctions) fused the Dβ1 12RSS to other DSBs along the 1.8-MB cassette-containing domain with a distribution similar to that of Dβ1 downstream coding-end joins in the ATM-proficient c-Myc-DJβ pro-B line and primary pro-B cells (Figures S1A–S1C; Table S1). Notably, there was increased but still low levels of translocations to $Ig\kappa$ (~0.2% of total junctions, Figure S1A) as compared to ATM-proficient pro-B cells. ATM-deficient BM c-Myc-DJβ pro-B cells also had similar patterns of Dβ1 23RSS coding-end junctions to those of ATM-proficient pro-B lines, except that they had

low-level translocations to *IgH* (∼0.07%) and *TCRα/δ* (∼0.05%) (Figures S1A–S1C). Finally, an ATM-deficient *v-Abl c-Myc*-DJβ pro-B line in which both cassette alleles were in the DJβ configuration generated few junctions, confirming that single 12RSS-containing alleles are inert (Figure S1E).

Abundant DSBs across the 1.8-Mb *c-Myc*-DJβ Loop Domain

We investigated the orientation of the thousands of Dβ1 downstream coding-end junctions within the 1.8-Mb *c-Myc* region in the *c-Myc*-DJβ *v-Abl* pro-B cell line. Junctions are denoted as in "+" orientation if prey sequence reads in a centromere-to-telomere direction and in "−" orientation if prey reads in the opposite direction (Chiarle et al., 2011). As the *c-Myc* E1 primer is centromeric to the bait Dβ1 downstream coding end, it captures junctions resulting in upstream excision circles and downstream deletions as + events and captures inversional junctions either upstream or downstream as − events (Figure 1D). Dβ1 downstream coding-end junctions near (within 5 kb) *c-Myc* occurred at similar frequency in + and − orientations (Figure S1F); strikingly, however, ∼95% of junctions to sequences further downstream of *c-Myc* occurred in deletional (+) orientation (Figure 1E). Similar results also were obtained with ATM-deficient *c-Myc*-DJβ pro-B cell lines (Figure S1C), even though their junctions involved Dβ1-12RSS ends (Figure S1D).

To gain insight into the basis for the well-defined boundaries of the DSB hotspot region flanking the *c-Myc*-DJβ cassette, we analyzed existing ChIP-seq data from BM pro-B cells (Lin et al., 2012) and found a cluster of CTCF and cohesin subunit Rad21-binding sites on both boundaries of this 1.8-Mb domain (Figure 1F). Moreover, the two clusters of CBEs were in convergent orientation (Figure 1F). Indeed, recent high-resolution in situ Hi-C data performed in mouse CH12-LX B cell lines (Rao et al., 2014) confirmed that this 1.8-Mb region is a well-defined convergent CBE-based loop domain that contains within it a strong 840-kb sub-loop that also extends to a convergent CBE (Figures 1F and 1G). HTGTS junction density within the 1.8-Mb domain in both ATM-proficient (Figures 1E–1G) and deficient *c-Myc*-DJβ pro-B cells (Figure S1C) correlated well with Hi-C interactions within the two loop domains.

RAG Generates DSBs in the 1.8-Mb *c-Myc*-DJβ Loop Domain

To test the relationship of frequent prey DSBs within the 1.8-Mb *c-Myc* loop domain to RAG-generated DSBs, we searched ATM-proficient and ATM-deficient *c-Myc*-DJβ junctions for sequence motifs in their vicinity. In this regard, the conserved 5′-CAC motif of the RSS heptamer is a position indicator for RAG cleavage, with cleavage invariably occurring 5′ to the CAC motif (Figure 1A). For convention, a CAC is considered in "forward" orientation if the presumed associated "coding" sequence is centromeric to the RSS and in "reverse" orientation if the presumed coding sequence is telomeric (Figure S2A). For widespread CACs, sequences in the coding position would not generally be gene segments; thus, we refer to them as surrogate coding ends. For analysis, we pooled and analyzed, respectively, all + junctions from the two *v-Abl* pro-B cell types and found that the majority occurred in putative surrogate coding sequences at or within

5 bp of a reverse CAC, with ∼30% joined directly to the surrogate coding sequence immediately flanking a CAC (Figures 2A, 2B, S2B, and S2C). There was no significant correlation with forward CACs (Figures 2C and S2D). These results suggest that the frequent DSBs within the 1.8-MB *c-Myc* domain occur at "cryptic RSSs" represented predominantly by a conserved CAC. Moreover, surrogate coding ends fused to the bait ends were processed similarly to normal coding ends during V(D)J recombination. The most highly recurrent hotspot DSBs within the 1.8-Mb domain tended to involve CACs within more canonical heptamers (Figure S2E). Finally, remarkably similar results were obtained from ATM-proficient and ATM-deficient *c-Myc*-DJβ BM pro-B cells (data not shown).

To unequivocally test the role of RAG in generating DSBs in the 1.8-Mb *c-Myc* loop domains, we deleted *Rag2* in the ATM-deficient *c-Myc*-DJβ pro-B cell line (Figure S2F). For HTGTS bait, we employed a Cas9/gRNA to generate DSBs 519 bp downstream of the *c-Myc*-DJβ cassette and designed a primer that allowed 5′ broken ends of these DSBs to be used as bait ("5′Cas9 bait ends"; Figure 2D). We then performed HTGTS on RAG-sufficient and RAG2-deficient G1-arrested pro-B cells. Recovered 5′Cas9 HTGTS junctions from RAG-sufficient ATM-deficient *c-Myc*-DJβ *v-Abl* pro-B cells correlated with reverse CACs in the 1.8-Mb domain as expected; however, unlike RAG-generated bait broken ends, the Cas9/gRNA-generated bait ends recovered junctions equally in + and − orientation (Figures 2D–2F and S2G). Performing these assays in ATM-deficient *v-Abl* pro-B cells that either lacked the *c-Myc*-DJβ cassette or were RAG2 deficient generated only a very few junctions within the 1.8-Mb domain, and these had no correlation with CACs (Figures 2D–2F and S2H). These findings confirm that RAG generates the off-target DSBs across the 1.8-Mb *c-Myc* domain in a *c-Myc*-DJβ-cassette-dependent fashion and also demonstrate that the asymmetric prey-joining preferences observed are specific to RAG-generated bait ends.

Paired Bona Fide RSSs Generate RAG Off-Target Activity in Loop Domains Genome-wide

We next tested whether other loop domains genome-wide could similarly be targets for such widespread RAG-generated DSBs if they contain bona fide RSS pairs. To insert bait RSSs into multiple genomic sites, we infected ATM-proficient and -deficient *v-Abl* pro-B lines with the pMX-DEL-SJ virus (referred to as "DEL-SJ"), which harbors a pair of divergent bona fide RSSs flanking an inverted GFP sequence (Figure 3A; Bredemeyer et al., 2006). V(D)J recombination between the divergent DEL-SJ RSSs fuses them in the chromosome and liberates the intervening GFP DNA within an excision circle generated via fusion of the surrogate coding ends (Figure 3A). We isolated six independent sub-clones from each genotype, each with a unique DEL-SJ-integration, treated them with STI-571, and generated HTGTS libraries with primers adjacent to either the construct 12RSS (12S primer) or 23RSS (23S primer) (Figure S3A). In all 12 DEL-SJ integration sites, the 12RSS and 23RSS junctions were confined within convergent CBE-based loop domains that ranged from 174 kb to 2.64 Mb in size (Table S2) and which often contained sub-domains flanked by convergent CBEs. For all integration sites, translocation junction density correlated
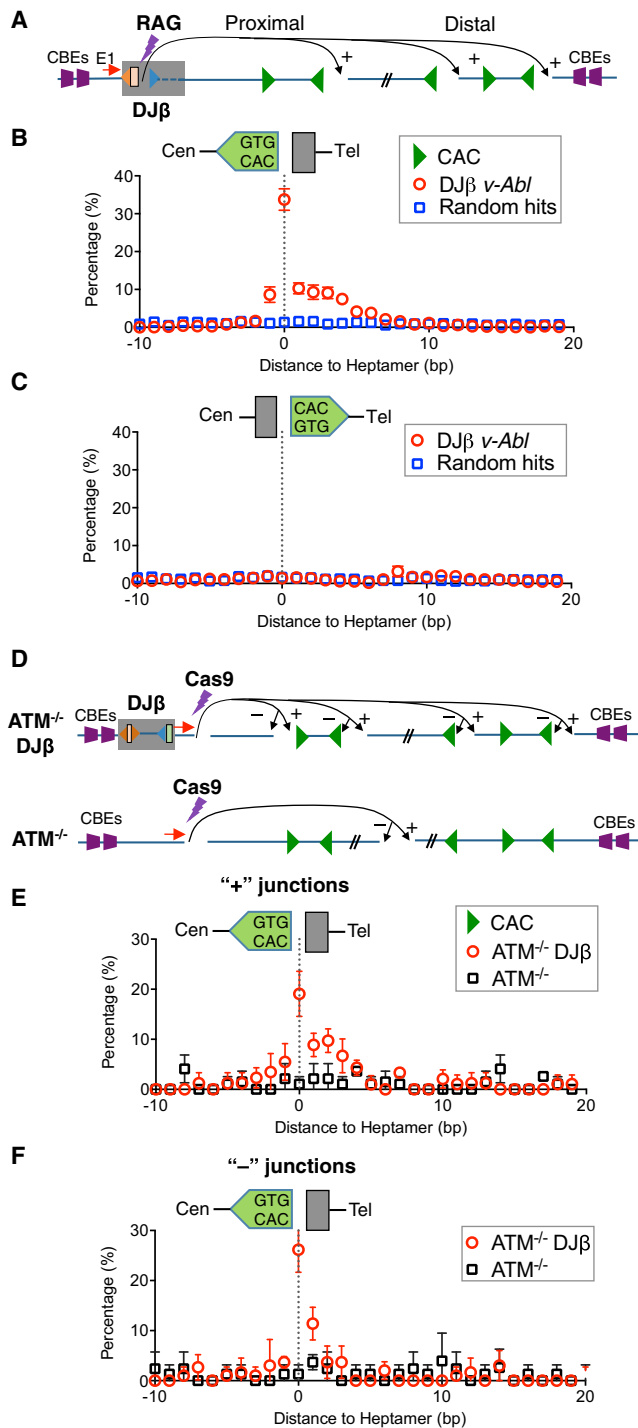
Cell

well with interaction intensities revealed by Hi-C. Representative findings from chromosome X, 4, 12, and 19 integrations are shown (Figures 3B–3K, S3B, and S3C). Notably, junctions detected from either 12RSS- or 23RSS-specific primers mostly occurred in deletional orientation independent of the orientation in which the DEL-SJ was integrated relative to the centromere (Figures 3B–3K, S3B, and S3C). As for the *c-Myc-DJβ* 1.8-Mb domain, hotspots also were apparent (Figures 3B–3K, S3B, and S3C).

We examined junction sequences within the two DEL-SJ loop domains on chromosome X and 4, respectively, for potential correlations with forward or reverse CACs. Deletional and excision circle junctions represented >95% of events for any given integration site (Figures 3B–3K). Strikingly, the vast majority of junctions were highly correlated with CACs; however, while bait RSSs joined to convergent upstream cryptic CACs, they joined to surrogate coding ends associated with downstream CACs in the same orientation to form apparent "hybrid" RSS-to-coding-end joins (Figures 4A–4C and S4A–S4C; but see below). Analysis of several other DEL-SJ-containing domains (on chromosomes 12 and 19) revealed precisely the same patterns despite diverse locations and relative chromosomal orientations (data not shown). Notably, upstream CACs were generally joined precisely to bait RSSs, but downstream joins to surrogate coding ends were often imprecise, with deletions of several nucleotides from the CAC border (Figures 4A–4C and S4A–S4C). The latter result, together with junctional sequence analysis of bait RSS ends (Figure S4D), indicates that RSS ends from the DEL-SJ construct that join downstream behave like surrogate coding ends in a V(D)J recombination-type of joining reaction.

Normal DEL-SJ V(D)J recombination generates fused RSS pairs at a high frequency (Figure 3A) that can be re-cleaved by RAG, with one cleavage product then being treated as an RSS end and the other as a surrogate coding end (Figure 4D; Meier and Lewis, 1993). Thus, the apparent downstream "hybrid joins" observed with the bait 12RSS, consistent with their end structure, could be generated from the fused intermediate. To test this possibility, we used as bait the 12RSSs of perfectly fused 12-23 joins of DEL-SJ within the chromosome X and 4 integrations, respectively. Indeed, this fused RSS pair faithfully recapitulated the joining patterns of the parental un-rearranged DEL-SJ construct in this location (Figures 4B, 4C, 4E, S4B, and S4E), demonstrating that the joining orientation of the two fused RSSs determines whether one or the other acts as an RSS end or surrogate coding end in the off-target V(D)J recombination joining reaction. Finally, we also generated HTGTS libraries from the surrogate coding ends (GFP primer) associated with 12RSS of DEL-SJ integrated into chromosome X. Such surrogate coding end junctions would not be re-cleaved by RAG. Correspondingly, nearly 90% of the 12RSS-associated coding ends joined downstream of the GFP primer to surrogate

Figure 2. RAG Generates DSBs across the 1.8-Mb *c-Myc*-DJβ Loop Domain

(A) Schematic of translocations between Dβ1 downstream coding ends and cryptic RSSs mostly represented by CAC motifs in the 1.8-Mb *c-Myc* domain in *c-Myc*-DJβ *vAbl* pro-B cells.

(B and C) Distance of Dβ1 downstream coding-end junctions to reverse CACs (B) or forward CACs (C) within the 1.8-Mb domain in *c-Myc*-DJβ *vAbl* pro-B cells. Direct joining to the nucleotide immediately adjacent to CAC is defined as 0 in this and following panels. Mean ± SD, n = 3.

(D) Schematic of translocations between Cas9/gRNA-initiated bait DSBs and DSBs in the *c-Myc* domain in ATM-deficient pro-B cells with (top) or without (bottom) the *c-Myc*-DJβ cassette.

(E and F) Distance of 5′Cas9 junctions in + (E) or − (F) orientation to reverse CACs in the *c-Myc* domain in cells defined in (D). Means ± SD, n = 3.

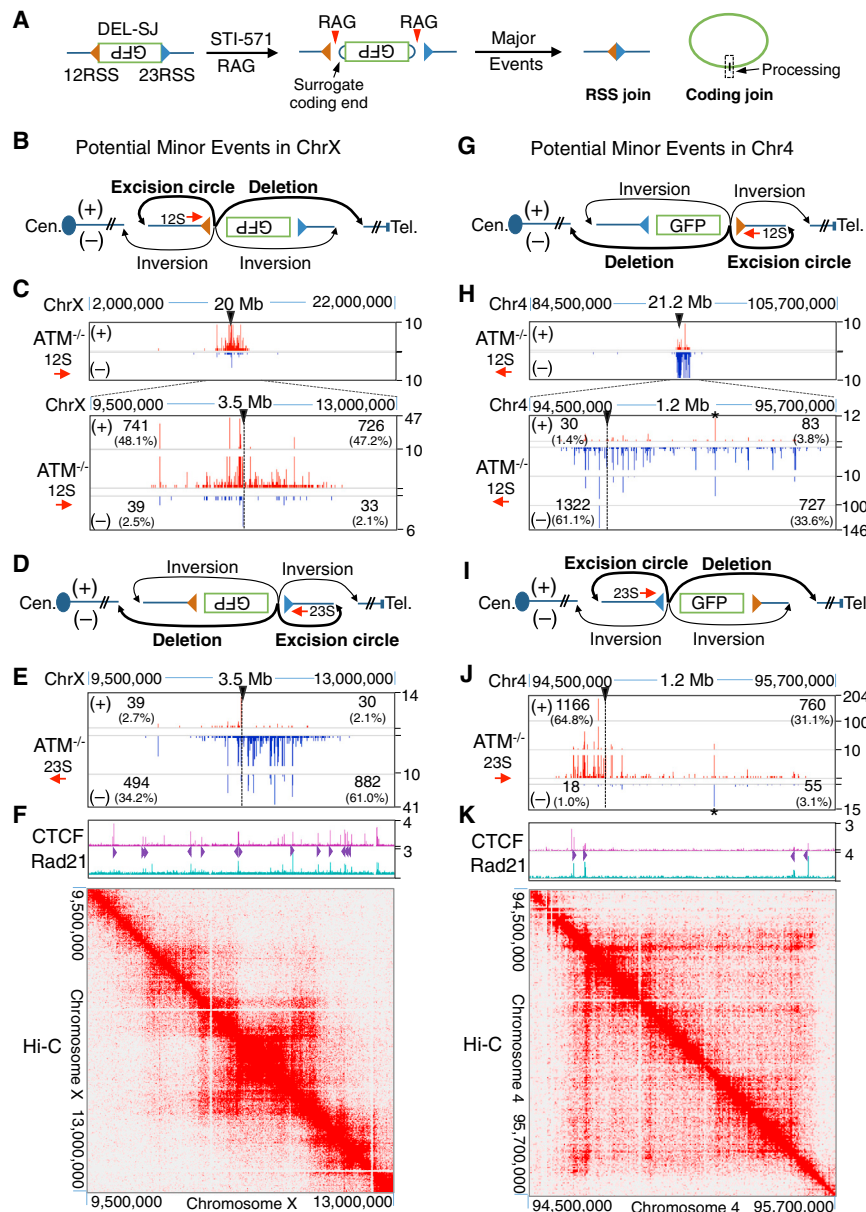See also Figure S2.

Cell

Figure 3. DSBs Restricted in Genome-wide DEL-SJ-Containing Loop Domains

(A) Diagram showing major RAG-initiated joins in DEL-SJ-containing *v-Abl* pro-B cells.

(B) Potential junctional outcomes between bait 12RSS and DSBs in *cis* include deletions, excision circles, and inversions.

(C) Profiles of 12RSS junctions within chromosome X in ATM-deficient *vAbl* pro-B cells. Black triangles indicate insertion site of DEL-SJ. Junction numbers and percentages in + or − orientation upstream or downstream of bait 12RSS are shown separately.

(D and E) Profiles of 23RSS junctions in the indicated 3.5-Mb region containing DEL-SJ on chromosome X.

(F) ChIP-seq profiles of CTCF/Rad21 (top) and heatmap of in situ Hi-C (bottom) in this 3.5-Mb region.

(G–K) 12RSS and 23RSS junctions across the 1-Mb DEL-SJ-containing loop domain on chromosome 4. Stars indicate stronger cryptic RSSs.

See also Figure S3 and Table S2.

coding ends adjacent to CACs (Figures 4F–4H). Together, our findings show that, for bona fide RSS pairs within a loop domain, both the RSS and the associated coding sequence join to convergent cryptic RSSs (CACs) and associated surrogate coding ends within a loop domain via a V(D)J recombination-like reaction.

Robust Detection of RAG Off-Targets Genome-wide Outside of Chromatin Domains

We further analyzed the 12RSS-associated coding end (GFP)-primed DEL-SJ HTGTS libraries from the X chromosome integration in ATM-deficient *v-Abl* lines and additional libraries from an integration on chromosome 1 in a different ATM-deficient *v-Abl* line (Figure S5A). Beyond the expected joining patterns within

the DEL-SJ-containing loop domains (Figures S5B–S5D), these libraries also revealed 107 translocation hotspots across the genome that all occurred at or near heptamers related to the canonical CACAGTG motif (Figures 5A, 5B, and S5E; Table S3). Notably, 60 of the 107 identified cryptic RSSs occurred in pairs in convergent orientation within <100 kb in the same domain (Figures 5A, 5C, and S5E; Table S3). HTGTS employing a primer upstream of the cryptic RSS in one such pair on chromosome 1 (Figure 5C) revealed thousands of deletional junctions involving two cryptic RSSs (Figures S5F–S5H). We compared locations of these 107 cryptic RSSs with existing pro-B H3K4me3 ChIP-seq data, which marks promoters, or H3K27Ac data that marks promoters and enhancers (Lane et al., 2014; Whyte et al., 2013). Strikingly, 97 of the 107 RAG off-targets overlapped with H3K27Ac-marked regions, with 38 overlapping with super-enhancers and 59 with typical enhancers. Of these, 65 overlapped with regions marked by both H3K4me3 and H3K27Ac (Figures 5D and 5E). These remarkably high correlations demonstrate that accessibility, beyond RAG binding, also is important for efficient RAG cleavage at cryptic RSSs.

IgH Employs CBE-Based Subdomains to Regulate RAG On- and Off-Target Activity

We applied HTGTS to test whether RAG on- and off-target activity in *IgH* is confined within IGCR1 CBE-based domains (Figure 6A). We employed an ATM-deficient *v-Abl* pro-B cell line that harbors a DFL16.1-J_H3 rearrangement, providing a
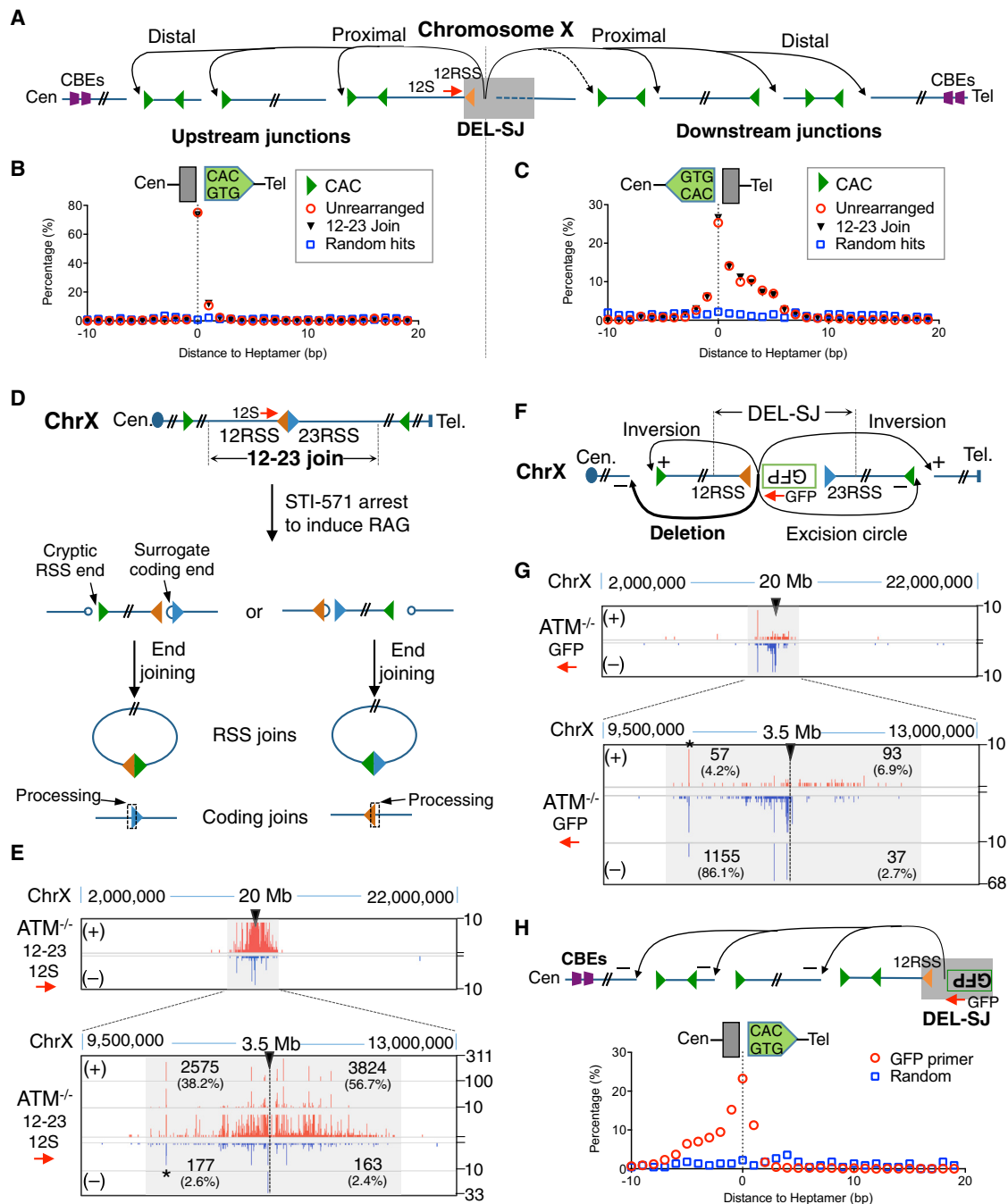
Figure 4. Orientation-Biased Joining of RAG-Initiated DSBs in Loop Domains

(A) Schematic of translocation between bait 12RSS and CACs within the DEL-SJ-containing loop domain on chromosome X in ATM-deficient *v-Abl* pro-B cells.

(B) Distance of 12RSS junctions upstream of bait 12RSS to forward CACs; no correlation was found with reverse CACs.

(C) Distance of 12RSS junctions downstream of bait 12RSS to reverse CACs; no correlation was found with forward CACs.

(D) Diagram of recombination output generated by RAG re-cleavage at perfect 12-23RSS joins.

(E) Profiles of 12RSS junctions of 12-23RSS join within chromosome X in ATM-deficient *vAbl* pro-B cells. Star indicates a relatively stronger cryptic RSS, and loop domain is shadowed in gray, also in (G).

(F) Diagram of bait surrogate coding ends associated with DEL-SJ 12RSS and the potential outcomes.

(G) Profiles of GFP primer junctions within chromosome X in ATM-deficient *vAbl* pro-B cells.

(H) Distance of GFP primer junctions downstream of bait surrogate coding ends to forward CACs.

See also Figure S4.
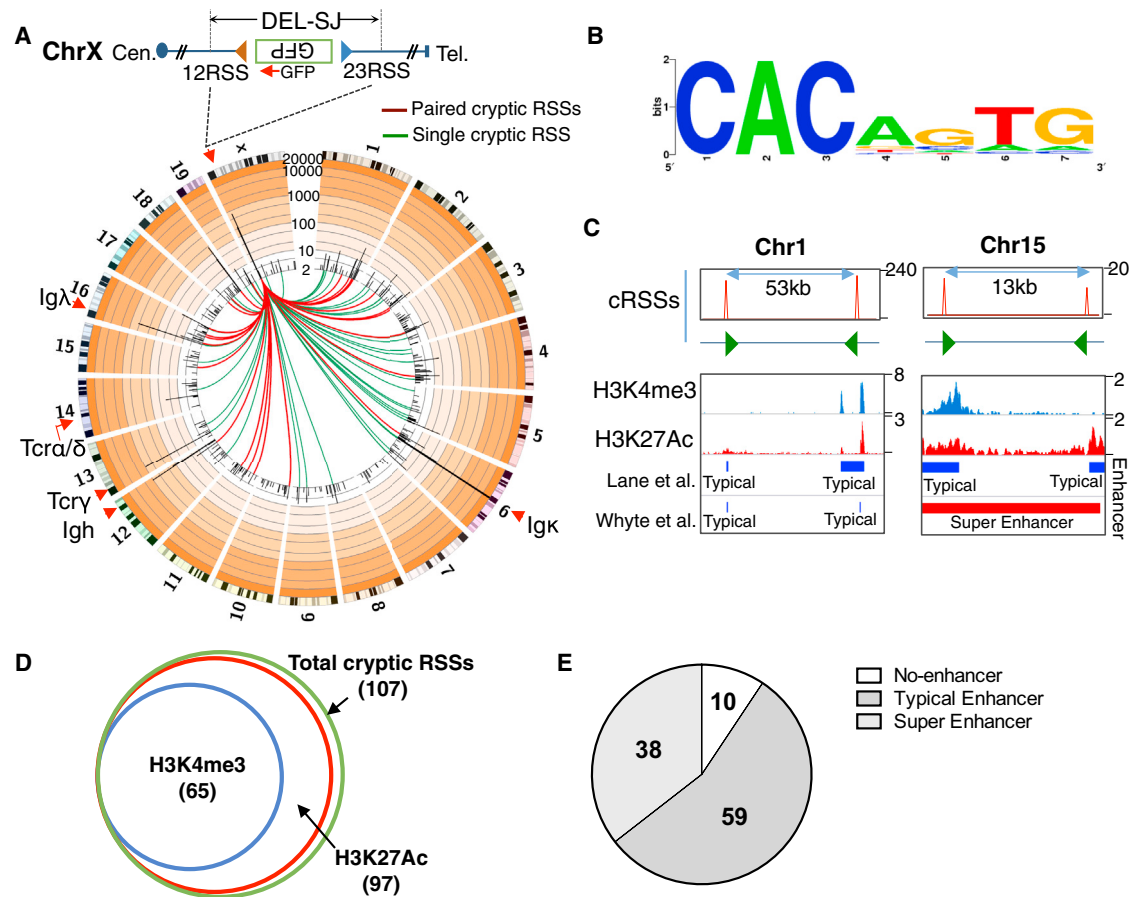
Cell

Figure 5. Genome-wide RAG Off-Targets

(A) Circos plot on a custom log scale showing genome-wide translocation profile of GFP primer junctions in DEL-SJ-integrated ATM-deficient *v-Abl* pro-B cells. Bin size is 2 Mb, and 50,000 HTGTS junctions are shown. Colored lines link break site to identified cryptic RSSs.

(B) Consensus sequence of cryptic RSS heptamer extrapolated from the 107 identified cryptic RSSs.

(C) Paired cryptic RSSs on chromosome 1 (left) and 15 (right). Green arrows indicate the position and orientation of cryptic RSSs in these four translocation hotspots.

(D) Venn diagram showing number of identified cryptic RSSs that overlap with H3K27Ac and H3K4me3.

(E) Pie chart showing number of identified cryptic RSSs that overlap with typical enhancers and super-enhancers.

See also Figure S5 and Table S3.

population of cells harboring a 5′D 12RSS expected to join to accessible upstream V_Hs 23RSSs (Alt et al., 2013). We used an HTGTS primer 82 bp upstream of the 5′ DFL16.1 12RSS to capture joins involving bait 5′DFL16.1-J_H3 RSS ends (Figure 6B). The majority of ~27,000 recovered *IgH* HTGTS junctions were on-targets at *IgH* bona fide RSSs (85%; Table S4), with most fusing the DFL16.1 5′RSS to a V_H 23RSS in physiologic (excision circle) orientation (Figures 6B and 6C). While such junctions involved multiple V_Hs across the 2.4-Mb V_H domain, they were biased toward proximal V_Hs, particularly V_H81x (38% of on-targets) (Figure 6C). We also observed substantial-inversional (+) joining between the DFL16.1 5′RSS and J_H4 23RSS (20% of *IgH* on-targets) (Figures 6B and 6C). Strikingly, IGCR1 deletion dramatically increased the number of DFL16.1 5′RSS junctions recovered (28-fold) (Figure S6A; Table S4), largely from markedly increased utilization of proximal V_Hs (48-fold) and, in particular, V_H81x (92% of junctions) (Figures 6D, S6B, and S6D). Corre-

spondingly, there was an 18-fold decrease in distal/middle V_H utilization and a 20-fold decrease in J_H4 junctions (Figures 6D, S6C, and S6F).

These *IgH* HTGTS studies also revealed low but highly reproducible off-target joining of DFL16.1 12RSS ends to DSBs within *IgH* that correlated with CACs (Figures S6G and S6H). Strikingly, ~95% of the off-target *IgH* junctions were within a tightly focused 12.3-kb region that contains the DFL16.1-J_H3 and is bounded upstream by IGCR1 and downstream by iEμ/Sμ (Figures 6E and 6F). We refer to this region as the iEμ/Sμ-to-IGCR1 "recombination domain." Strikingly, deletion of IGCR1 from this ATM-deficient pro-B line dramatically changed the profile of off-target DSBs, permitting them to spread ~120 kb upstream into the proximal V_Hs while decreasing the percentage of off-target junctions in the former iEμ/Sμ to IGCR1 domain to 13% (Figures 6G and S6I). Thus, IGCR1 deletion established a new iEμ/Sμ-to-proximal-V_H recombination domain in which RAG activity on

both cryptic RSSs and proximal V_Hs bona fide RSSs is re-focused. As in other domains, RAG off-target activity was highly dependent on convergent CAC orientation once several kb from the DFL16.1 5'RSS break site (Figures S6H and S6J).

DISCUSSION

Mechanism of RAG Off-Target Activity

We report a major form of RAG off-target activity that eluded prior investigations. Remarkably, this activity is largely confined to loop domains containing paired bona fide RSSs, with cleavage requiring only recognition of a simple CAC motif. Also remarkable, this off-target RAG activity is directionally oriented such that RSS ends from paired bona fide RSSs join to convergent CAC-containing motifs, while coding ends from paired bona fide RSSs join to surrogate coding ends associated with a CAC. Thus, RSSs and corresponding coding ends join with the same patterns and in the same locations, consistent with V(D)J recombination (Figure 4). Such orientation dependence is most readily explained by a linear tracking mechanism (Yancopoulos et al., 1984; Figure 7). Based on RAG structural information (Kim et al., 2015), we propose a working model (Figures 7 and S7). This model assumes that formation and activation of the tetrameric RAG1/2 complex (Lapkouski et al., 2015) requires binding of paired bona fide RSSs (Figure 7B). We hypothesize that one or the other of these occasionally escapes the activated complex, allowing cryptic RSSs to replace them. The replacement process could involve diffusion of proximal cryptic RSSs or unidirectional tracking to more distal cryptic RSSs downstream. Once appropriately positioned in the activated complex, cryptic RSSs and surrogate coding ends could, likely at reduced frequency, be cleaved and joined to their remaining bona fide counterpart via a reaction that preserves most aspects of normal V(D)J recombination (Figure 7G). This general tracking model explains all aspects of our findings, including tracking from the two bona fide RSSs of a pair in opposite directions around the loop.

Implications for Normal Loop Domain Functions

An obvious and important question arises as to why RAG activity is so highly restricted within loop domains containing the initiating paired bona fide RSSs. One contributing factor could be high interaction frequency of DNA in chromatin across these domains (Alt et al., 2013). In this regard, DSB ends find and join to ends of other DSBs within such domains at higher frequency than elsewhere in the genome (Alt et al., 2013; Zhang et al., 2012). During IgH CSR, this phenomenon promotes proper and frequent joining of AID-initiated DSBs (Dong et al., 2015; Zarrin et al., 2007). Such DSB interactions are evident in our current studies in which Cas9/gRNA-generated DSBs frequently join to RAG off-target DSBs within the same loop. However, distinct from RAG-generated RSS or coding ends, a given Cas9/gRNA bait end joins to both cryptic RSS ends and surrogate coding ends of RAG off-target DSBs. Another apparent difference is that site-specific nuclease- or AID-generated DSBs appear to find off-targets in other regions across the genome much more readily than do RAG-generated DSBs, even in WT cells (Chiarle et al., 2011; Dong et al., 2015; Frock et al., 2015). The almost exclusive restriction of RAG off-targets to paired bona fide

RSS-containing loops implies that an additional mechanism enforces such RAG activity.

The tracking mechanism can explain the additional restriction of RAG activity within a given loop (Figure 7). We do not know the mechanism that propels RAG tracking, although transcription and/or cohesin might be involved (e.g., Nichols and Corces, 2015). However, it is reasonable to assume that tracking is terminated when it encounters a block imposed by the CTCF/cohesin-bound convergent CBE pair or similar loop-forming interactions (Figure 7F). Such blockage would terminate tracking in each direction from paired bona fide RSSs and limit off-target RAG activity to the loop. In support of this model, deletion of the CBE-based IGCR1 allows RAG off-target activity to extend from its initial highly restricted location in the D-J_H recombination domain to >100 kb upstream, where new boundaries may form via V_H CBEs and/or associated factors (Figure 6G). Beyond regulating V(D)J recombination, related loop domain functions might impact on other activities constrained within them, including replication (Pope et al., 2014) and promoter/enhancer interactions (Dowen et al., 2014).

IgH Locus Regulation

Regulated IgH V_H-to-DJ_H recombination depends on the integrity of the two divergent CBEs within IGCR1, likely via formation of loop domains that focus RAG activity on D_Hs and J_Hs (Guo et al., 2011; Lin et al., 2015). Our HTGTS studies provide additional insights into IgH V(D)J recombination regulation (Figure 6). In a DJ_H-rearranged pro-B cell line, on-target rearrangements of the 5'D RSS occur to RSSs of V_Hs across the locus but predominantly to 3' V_Hs (Figure 6C). As most RAG off-target activity is focused in a small 12.3-kb recombination domain from IGCR1 to the i$E\mu$/Sμ boundary (Figure S6G), the recombination domain in these cells does not extend downstream to 3'CBEs as perhaps anticipated. This restriction could be due to IGCR1 CBE looping with non-CBE elements at i$E\mu$/Sμ (Guo et al., 2011) and/or by tracking limitations imposed by a unidirectional mechanism. In the DJ_H-rearranged cells, bona fide V(D)J recombination at upstream V_Hs in the absence of corresponding off-target activity, even in proximal portions of the locus, is consistent with V_Hs entering the recombination domain by a specialized mechanism operating subsequent to locus contraction (Bossen et al., 2012). Based on off-target activity as an assay, IGCR1 deletion extends the recombination domain linearly into proximal V_Hs, resulting in a huge overall V(D)J recombination increase, involving V_H81x and other proximal V_Hs (Figures 6D and S6D). This increase may be facilitated by increased interaction frequency gained by placing V_H 23RSSs in the same loop domain as the 5' DFL16.1 12RSS and/or by a tracking contribution. Finally, a unidirectional RAG tracking mechanism also might explain why 3'D 12RSSs, but not 5'D 12RSSs, are used developmentally in D-to-J_H rearrangements.

RAG Off-Target Activity, Chromosomal Rearrangements, and Cancer

We prove our hypothesis that inserting paired bona fide RSSs into c-Myc activates RAG-generated DSBs at cryptic RSSs over a long region downstream that, in the context of ATM deficiency, promotes oncogenic translocations. These findings explain how paired bona fide RSSs within a Tcrα excision circle
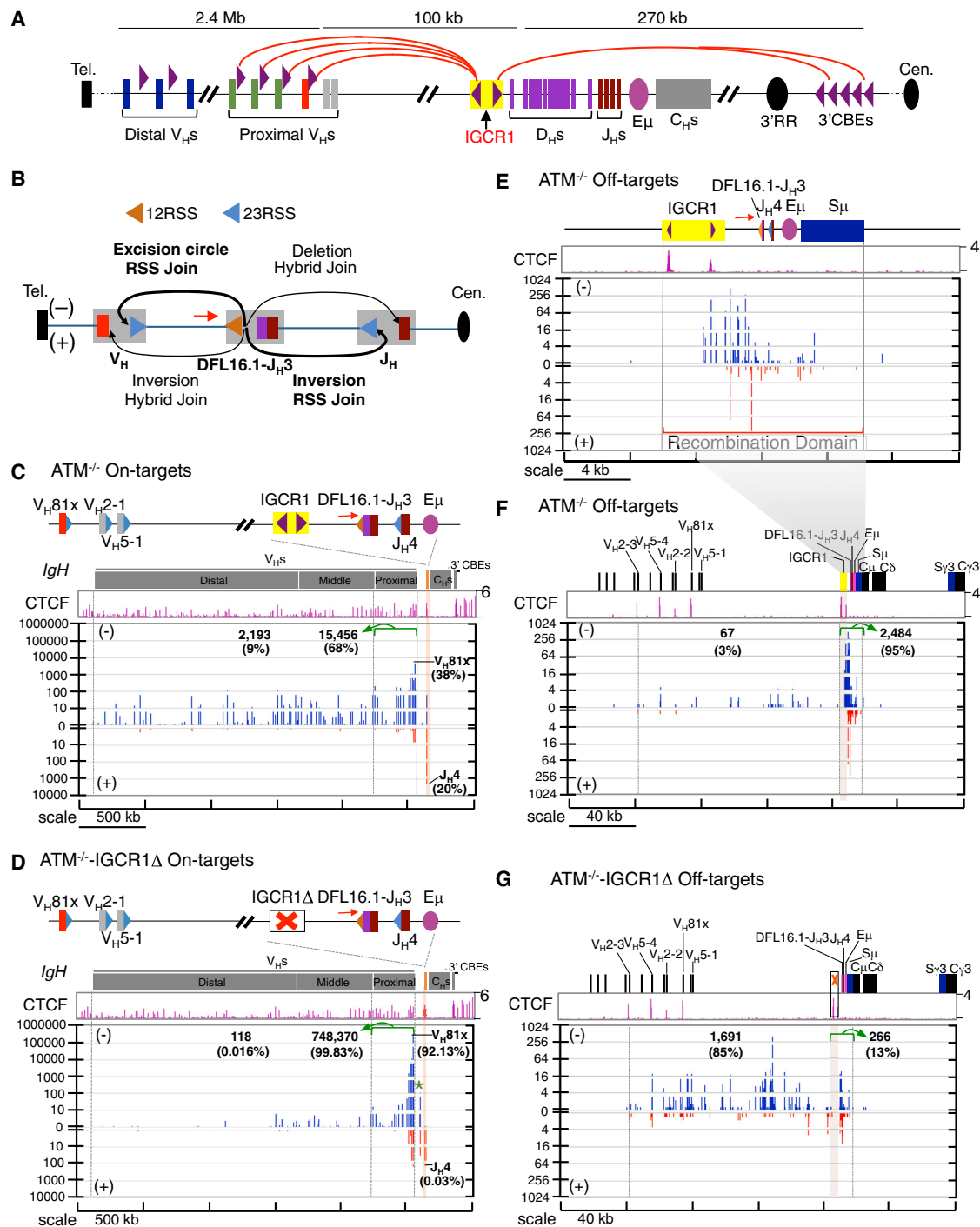
Cell

Figure 6. Effect of IGCR1 Deletion on RAG Targeting in the *IgH* Locus

(A) Schematic of murine *IgH* locus. Purple arrowheads indicate position and orientation of CBEs. Red arches show proposed looping between convergent *IgH* CBEs (Lin et al., 2015).

(B) Illustration of possible joining outcomes between bona fide RSSs from DFL16.1 5′ RSS bait end to upstream V_H and downstream J_H broken ends. Red arrows in all panels indicate the position and orientation of HTGTS primer.

(C and D) Top panels are *IgH* annotation track. Middle panels show CTCF ChIP-seq profiles. Bottom panels are pooled HTGTS junction profiles for *IgH* bona fide RSSs (n = 3). Junctions are displayed as stacked tracks (log scale between tracks, linear scale within each track). Junction numbers and percentages as of total *IgH* on-targets in indicated regions are shown. We note that, beyond junctions described in the text, we also detected junctions between DFL16.1 5′RSS and pseudo D_H RSSs in the V_H-to-D_H intervening region in ATM$^{-/-}$-IGCR1Δ cells (green star; see also Figure S6E). We also found a very small number of hybrid joins to V_H coding ends (+ joins) or J_H4 coding ends (− joins).

(legend continued on next page)
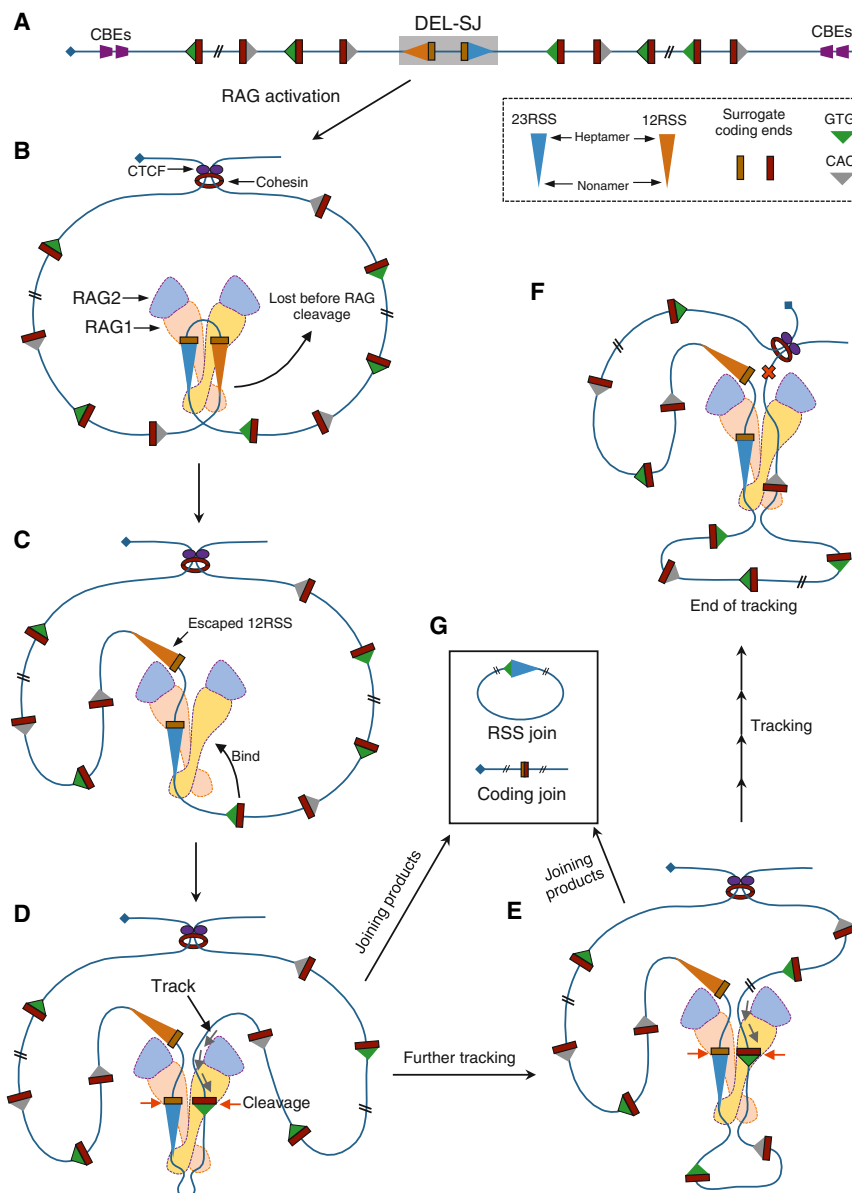
Figure 7. Linear Tracking Model for Orientation-Biased Usage of CACs in the Paired Bona Fide RSSs-Containing Loop Domains

(A) Linear map of a DEL-SJ-containing loop domain.

(B) Activation of RAG by paired bona fide RSSs in the CTCF/cohesin-anchored loop domain.

(C) One bona fide RSS may escape at a very low frequency leaving the activated complex.

(D) A CAC in convergent orientation of the remaining bona fide RSS binds to RAG to facilitate cleavage or initiate unidirectional tracking.

(E) Further tracking resulting in usage of another convergent CAC.

(F) RAG tracking in the loop in either direction is stopped by the boundary formed by the CTCF/cohesin complex.

(G) Joining products from pairing and cleavage between remaining bona fide RSS and convergent CACs.

More details are in the text. Various modifications of this basic tracking model can be conceived to explain joining patterns of proximal V, D, and J segments in endogenous antigen receptor loci (data not shown).

See also Figure S7.

Thus, ATM limits potential RAG-initiated translocations by promoting joining of RAG-initiated DSBs at RSSs and cryptic RSSs within a loop. Our findings also provide a mechanism for oncogenic translocations to sequences far downstream of *c-Myc* in C-NHEJ/p53 double-deficient pro-B cells (Alt et al., 2013). In this regard, we find cryptic RSSs in the *c-Myc* 1.8-Mb domain that are closer to consensus (Merelli et al., 2010) and, therefore, may drive RAG-initiated DSBs at other cryptic RSSs in this domain that become liberated from post-cleavage complexes in the absence of C-NHEJ.

We also found 107 genome-wide cryptic RSSs, not related to antigen receptor loci or paired bona fide RSSs-containing domains, that were DSB and translocation targets in ATM-deficient *v-Abl* pro-B cells (Figure 5; Table S3). This set of cryptic RSSs tended to have heptamers even closer to consensus than recurrent hotspots within paired bona fide RSSs-containing loops (Figure 5B versus S2E). Many of these translocation target RSSs occurred in pairs separated by <100 kb (Table S3), with each member of the pair falling directly within enhancer and/or promoter regions (Figures 5C–5E). Enhancer/promoter loops also might increase the frequency

fragment integrated into the *HPRT* locus in leukemia cells causes further genomic aberrations (Messier et al., 2006) and also support the hypothesis that translocations downstream of *c-Myc* in human B cell lymphomas involve cryptic RSSs (Kroenlein et al., 2012). Given that cryptic RSS targeting downstream of *c-Myc* occurs in both WT and ATM-deficient pro-B cells, one role of ATM in suppressing such translocations would be through stabilizing ends in RAG post-cleavage complexes to facilitate their joining via V(D)J recombination (Bredemeyer et al., 2006).

(E) Pooled RAG off-target junction profile of a 24-kb region including the recombination domain for ATM-deficient cells (n = 3).

(F and G) Pooled RAG off-target junction profiles of a 240-kb region including the recombination domain for ATM-deficient cells with (F) or without (G) IGCR1, respectively (n = 3). Brown shadowed region marks the location of IGCR1. Junction numbers and percentages as of total *IgH* off-targets in indicated regions are shown.

See also Figure S6 and Table S4.

Cell

with which such paired cryptic RSSs are juxtaposed to form stable RAG synaptic complexes. Strikingly, all 30 pairs of these cryptic RSS translocation targets were in convergent orientation (Table S3), similar to most proximal paired bona fide RSSs within antigen receptor loci (Bossen et al., 2012) and the majority of cryptic RSSs captured by bona fide RSSs within loop domains. Thus, to serve as a strong genome-wide translocation target, cryptic RSS require a good heptamer, location in enhancers and/or promoters, and convergent pairing with another good cryptic RSS in the same loop. Finally, our findings provide a mechanistic basis for recurrent oncogenic chromosomal interstitial deletions in tumors arising from developing human lymphocytes (Larmonie et al., 2013; Mullighan et al., 2008; Papaemmanuil et al., 2014).

EXPERIMENTAL PROCEDURES

Cell Lines

BM pro-B cells were purified by αB220 selection from ATM-proficient and -deficient c-Myc-DJβ mice (Tepsuporn et al., 2014) and were cultured in opti-MEM medium with 10% (v/v) FBS plus IL-7 (2 ng/ml) and SCF (2 ng/ml) for 4 days. The v-Abl pro-B cells were cultured in RPMI medium with 15% (v/v) FBS; cells were treated with STI-571 (3 μM) for 4 days to express RAG. WT and ATM-deficient v-Abl pro-B cell lines were described previously (Zha et al., 2011). ATM-proficient and -deficient c-Myc-DJβ v-Abl pro-B cell lines were made specifically for this study from Eμ-Bcl-2 transgenic mice of the corresponding genotypes. We included the Eμ-Bcl-2 transgene in these cells to protect STI-571-treated (G1-arrested) v-Abl pro-B cells from apoptosis; prior work showed that Bcl-2 expression has no effect on V(D)J recombination (Zha et al., 2011).

RAG On- and Off-Targets

HTGTS was performed and analyzed as previously described with modifications (Frock et al., 2015). Primers for HTGTS are listed in Table S5. Due to the very low junctional diversity of bona fide V(D)J recombination RSS joins and coding joins, we included duplicate junctions in our analyses of G1-arrested v-Abl cells to better reflect the actual frequencies of the various classes of bona fide and off-target junctions. Where approximate percentage and/or numbers of different classes of junctions are indicated (e.g., c-Myc-DJβ or IgH), we controlled for reproducibility by performing at least three independent experiments (e.g., Table S4). RAG off-target hotspots were identified by MACS2 (Zhang et al., 2008), with extend size (extsize) at 20 bp and false discovery rate (FDR) cut-off at 10^{-9}. See Supplemental Information for more details.

ChIP-Seq and Hi-C Data

CTCF and Rad21 ChIP-seq data were extracted from Lin et al. (2012) (GEO: GSE40173); H3K4me3 and H3K27Ac ChIP-seq data were extracted from Lane et al. (2014) (GEO: GSE48555). These data are from BM pro-B cells. We re-analyzed ChIP-seq data with Chilin software (http://cistrome.org/chilin/) in the simple model against mm9. Enhancer annotation was either extracted directly from Whyte et al. (2013) (GEO: GSE44288) or identified by Homer software (Heinz et al., 2010) from re-analyzed H3K27Ac ChIP-seq data (Lane et al., 2014). In situ Hi-C data for CH12-LX B cells was extracted and displayed (KR normalization) by Juicebox software (Rao et al., 2014).

ACCESSION NUMBERS

The Gene Expression Omnibus (GEO) accession number for the datasets reported in this paper is GEO: GSE73007.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.016.

REFERENCES

Alt, F.W., Zhang, Y., Meng, F.-L., Guo, C., and Schwer, B. (2013). Mechanisms of programmed DNA lesions and genomic instability in the immune system. Cell 152, 417–429.

Boehm, T., Mengle-Gaw, L., Kees, U.R., Spurr, N., Lavenir, I., Forster, A., and Rabbitts, T.H. (1989). Alternating purine-pyrimidine tracts may promote chromosomal translocations seen in a variety of human lymphoid tumours. EMBO J. 8, 2621–2631.

Bossen, C., Mansson, R., and Murre, C. (2012). Chromatin topology and the regulation of antigen receptor assembly. Annu. Rev. Immunol. 30, 337–356.

Bredemeyer, A.L., Sharma, G.G., Huang, C.-Y., Helmink, B.A., Walker, L.M., Khor, K.C., Nuskey, B., Sullivan, K.E., Pandita, T.K., Bassing, C.H., and Sleckman, B.P. (2006). ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. Nature 442, 466–470.

Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.-J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., et al. (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell 147, 107–119.

Degner, S.C., Wong, T.P., Jankevicius, G., and Feeney, A.J. (2009). Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. J. Immunol. 182, 44–48.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380.

Dong, J., Panchakshari, R.A., Zhang, T., Zhang, Y., Hu, J., Volpi, S.A., Meyers, R.M., Ho, Y.-J., Du, Z., Robbiani, D.F., et al. (2015). Orientation-specific joining of AID-initiated DNA breaks promotes antibody class switching. Nature 525, 134–139.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K., and Young, R.A. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell 159, 374–387.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.-J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat. Biotechnol. 33, 179–186.

Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.-L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., et al. (2011). CTCF-binding elements mediate control of V(D)J recombination. Nature 477, 424–430.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589.

Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R., and Schatz, D.G. (2010). The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. Cell 141, 419–431.

Kim, M.-S., Lapkouski, M., Yang, W., and Gellert, M. (2015). Crystal structure of the V(D)J recombinase RAG1-RAG2. Nature 518, 507–511.

Kroenlein, H., Schwartz, S., Reinhardt, R., Rieder, H., Molkentin, M., Gökbu-get, N., Hoelzer, D., Thiel, E., and Burmeister, T. (2012). Molecular analysis of the t(2;8)/MYC-IGK translocation in high-grade lymphoma/leukemia by long-distance inverse PCR. Genes Chromosomes Cancer 51, 290–299.

Lane, A.A., Chapuy, B., Lin, C.Y., Tivey, T., Li, H., Townsend, E.C., van Bode-gom, D., Day, T.A., Wu, S.-C., Liu, H., et al. (2014). Triplication of a 21q22 region contributes to B cell transformation through HMGN1 overexpression and loss of histone H3 Lys27 trimethylation. Nat. Genet. 46, 618–623.

Lapkouski, M., Chuenchor, W., Kim, M.-S., Gellert, M., and Yang, W. (2015). Assembly Pathway and Characterization of the RAG1/2-DNA Paired and Signal-end Complexes. J. Biol. Chem. 290, 14618–14625.

Larmonie, N.S.D., Dik, W.A., Meijerink, J.P.P., Homminga, I., van Dongen, J.J.M., and Langerak, A.W. (2013). Breakpoint sites disclose the role of the V(D)J recombination machinery in the formation of T-cell receptor (TCR) and non-TCR associated aberrations in T-cell acute lymphoblastic leukemia. Hae-matologica 98, 1173–1184.

Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chan-dra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat. Immunol. 13, 1196–1204.

Lin, S.G., Guo, C., Su, A., Zhang, Y., and Alt, F.W. (2015). CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V(D)J recombination. Proc. Natl. Acad. Sci. USA 112, 1815–1820.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161, 1012–1025.

Meier, J.T., and Lewis, S.M. (1993). P nucleotides in V(D)J recombination: a fine-structure analysis. Mol. Cell. Biol. 13, 1078–1092.

Merelli, I., Guffanti, A., Fabbri, M., Cocito, A., Furia, L., Grazini, U., Bonnal, R.J., Milanesi, L., and McBlane, F. (2010). RSSsite: a reference database and pre-diction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes. Nucleic Acids Res. 38, W262–W267.

Messier, T.L., O'Neill, J.P., and Finette, B.A. (2006). V(D)J recombinase medi-ated inter-chromosomal HPRT alterations at cryptic recombination signal se-quences in peripheral human T cells. Hum. Mutat. 27, 829.

Mullighan, C.G., Miller, C.B., Radtke, I., Phillips, L.A., Dalton, J., Ma, J., White, D., Hughes, T.P., Le Beau, M.M., Pui, C.-H., et al. (2008). BCR-ABL1 lympho-blastic leukaemia is characterized by the deletion of Ikaros. Nature 453, 110–114.

Nakahashi, H., Kwon, K.-R.K., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A., et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. 3, 1678–1689.

Nichols, M.H., and Corces, V.G. (2015). A CTCF Code for 3D Genome Archi-tecture. Cell 162, 703–705.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partition-ing of the regulatory landscape of the X-inactivation centre. Nature 485, 381–385.

Ong, C.-T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. 15, 234–246.

Onozawa, M., and Aplan, P.D. (2012). Illegitimate V(D)J recombination involving nonantigen receptor loci in lymphoid malignancy. Genes Chromo-somes Cancer 51, 525–535.

Papaemmanuil, E., Rapado, I., Li, Y., Potter, N.E., Wedge, D.C., Tubio, J., Alexandrov, L.B., Van Loo, P., Cooke, S.L., Marshall, J., et al. (2014). RAG-mediated recombination is the predominant driver of oncogenic rearrange-ment in ETV6-RUNX1 acute lymphoblastic leukemia. Nat. Genet. 46, 116–125.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 153, 1281–1295.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., et al. (2014). Topologically associating do-mains are stable units of replication-timing regulation. Nature 515, 402–405.

Ranganath, S., Carpenter, A.C., Gleason, M., Shaw, A.C., Bassing, C.H., and Alt, F.W. (2008). Productive coupling of accessible Vbeta14 segments and DJbeta complexes determines the frequency of Vbeta14 rearrangement. J. Immunol. 180, 2339–2346.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals prin-ciples of chromatin looping. Cell 159, 1665–1680.

Schatz, D.G., and Swanson, P.C. (2011). V(D)J recombination: mechanisms of initiation. Annu. Rev. Genet. 45, 167–202.

Teng, G., Maman, Y., Resch, W., Kim, M., Yamane, A., Qian, J., Kieffer-Kwon, K.-R., Mandal, M., Ji, Y., Meffre, E., et al. (2015). RAG represents a widespread threat to the lymphocyte genome. Cell 162, 751–765.

Tepsuporn, S., Hu, J., Gostissa, M., and Alt, F.W. (2014). Mechanisms that can promote peripheral B-cell lymphoma in ATM-deficient mice. Cancer Immunol. Res. 2, 857–866.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evo-lution of chromosomal domain architecture. Cell Rep. 10, 1297–1309.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319.

Yancopoulos, G.D., and Alt, F.W. (1986). Regulation of the assembly and expression of variable-region genes. Annu. Rev. Immunol. 4, 339–368.

Yancopoulos, G.D., Desiderio, S.V., Paskind, M., Kearney, J.F., Baltimore, D., and Alt, F.W. (1984). Preferential utilization of the most JH-proximal VH gene segments in pre-B-cell lines. Nature 311, 727–733.

Zarrin, A.A., Del Vecchio, C., Tseng, E., Gleason, M., Zarin, P., Tian, M., and Alt, F.W. (2007). Antibody class switching mediated by yeast endonuclease-generated DNA breaks. Science 315, 377–381.

Zha, S., Guo, C., Boboila, C., Oksenych, V., Cheng, H.-L., Zhang, Y., Wese-mann, D.R., Yuen, G., Patel, H., Goff, P.H., et al. (2011). ATM damage response and XLF repair factor are functionally redundant in joining DNA breaks. Nature 469, 250–254.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137.

Zhang, Y., McCord, R.P., Ho, Y.-J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell 148, 908–921.

Zuin, J., Dixon, J.R., van der Reijden, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van IJcken, W.F.J., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. USA 111, 996–1001.

Cell

Estrogen Receptor β Modulates Apoptosis Complexes and the Inflammasome to Drive the Pathogenesis of Endometriosis

Graphical Abstract
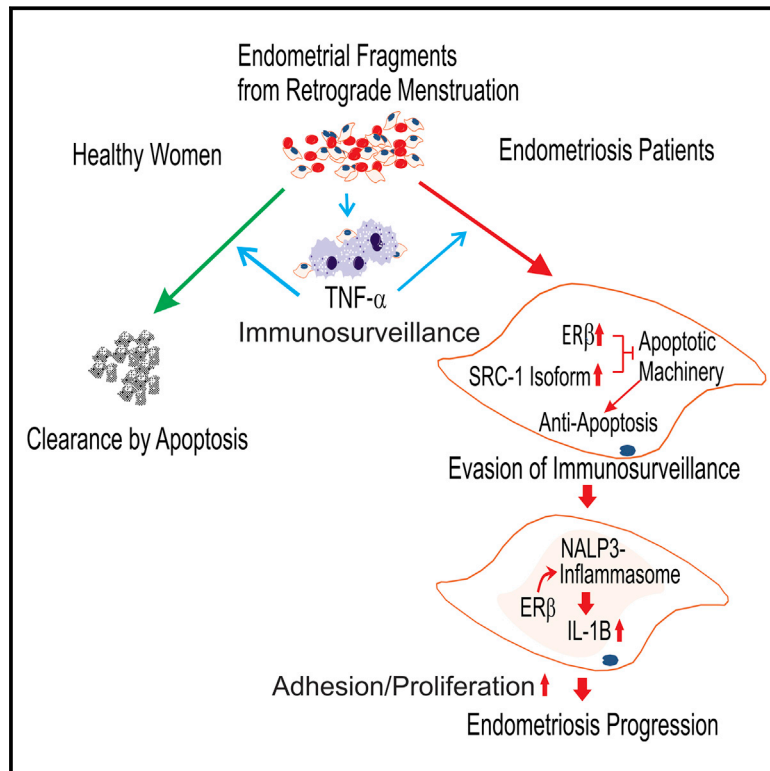
Authors

Sang Jun Han, Sung Yun Jung, San-Pin Wu, ..., Ming-Jer Tsai, Francesco J. DeMayo, Bert W. O'Malley

Correspondence

berto@bcm.edu

In Brief

Estrogen receptor β (ERβ) plays a unique role in endometriotic tissue, where it interacts with the cytoplasmic apoptotic machinery and inflammasome complex to prevent TNF-α-induced cell death and enhance adhesion and proliferative activities of endometriotic tissues. This non-genomic action of ERβ has a predominant role in endometriosis progression.

Highlights

- ERβ function drives progression

- ERβ interacts with cytoplasmic apoptotic machinery to prevent TNF-α-induced apoptosis

- ERβ-inflammasome enhances adhesion and proliferates

- ERβ EMT signaling induces invasion in ectopic lesions

CrossMark

CellPress

Estrogen Receptor β Modulates Apoptosis Complexes and the Inflammasome to Drive the Pathogenesis of Endometriosis

Sang Jun Han,[1] Sung Yun Jung,[1,4] San-Pin Wu,[1] Shannon M. Hawkins,[2] Mi Jin Park,[1] Satoru Kyo,[3] Jun Qin,[1,4] John P. Lydon,[1] Sophia Y. Tsai,[1] Ming-Jer Tsai,[1] Francesco J. DeMayo,[1,5] and Bert W. O'Malley[1,5,*]

[1]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA
[2]Departments of Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX 77030, USA
[3]Department of Obstetrics and Gynecology, Kanazawa University, School of Medical Science, Ishikawa 920-8640, Japan
[4]Alkek Center for Molecular Discovery, Verna and Marrs McLean, Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA
[5]These authors contributed equally to this work
*Correspondence: berto@bcm.edu
 http://dx.doi.org/10.1016/j.cell.2015.10.034

SUMMARY

Alterations in estrogen-mediated cellular signaling play an essential role in the pathogenesis of endometriosis. In addition to higher estrogen receptor (ER) β levels, enhanced ERβ activity was detected in endometriotic tissues, and the inhibition of enhanced ERβ activity by an ERβ-selective antagonist suppressed mouse ectopic lesion growth. Notably, gain of ERβ function stimulated the progression of endometriosis. As a mechanism to evade endogenous immune surveillance for cell survival, ERβ interacts with cellular apoptotic machinery in the cytoplasm to inhibit TNF-α-induced apoptosis. ERβ also interacts with components of the cytoplasmic inflammasome to increase interleukin-1β and thus enhance its cellular adhesion and proliferation properties. Furthermore, this gain of ERβ function enhances epithelial-mesenchymal transition signaling, thereby increasing the invasion activity of endometriotic tissues for establishment of ectopic lesions. Collectively, we reveal how endometrial tissue generated by retrograde menstruation can escape immune surveillance and develop into sustained ectopic lesions via gain of ERβ function.

INTRODUCTION

Endometriosis is a medical condition in which endometrial cells are deposited and grow outside the uterine cavity (Bulun, 2009; Giudice, 2010). Severe symptoms of endometriosis are typically observed in 6%–10% of reproductive-aged women (Simoens et al., 2007). Among patients with endometriosis, approximately 50% have major pelvic pain, and 40%–50% have fertility problems (Eskenazi and Warner, 1997; Ozkan et al., 2008). In these patients, endometriosis-associated symptoms negatively impact their health and quality of life (Moradi et al., 2014).

To improve the efficiency of endometriosis therapies, it is important to dissect the unique molecular properties of endometriotic tissues compared with normal endometria. Previous studies identified several endocrine properties associated with endometriotic tissues. Altered estrogenic signaling pathways have been reported in endometriosis pathogenesis (Bulun, 2009). Endometriotic lesions have been reported to contain higher 17β-estradiol levels than normal endometria due to the elevated expression of 17β-hydroxysteroid dehydrogenase-1 and aromatase genes compared with normal endometria (Acién et al., 2007; Delvoux et al., 2009). These higher levels of local 17β-estradiol could play a role in the proliferation of endometriotic tissues (Zhang et al., 2010). This increased 17β-estradiol binds and activates estrogen receptors (ERs) in endometriotic tissues to stimulate estrogen-dependent growth. There are two different forms of the ER, usually referred to as α and β, each encoded by a different gene, *ESR1* and *ESR2*, respectively. Prior studies with ERα$^{-/-}$, ERβ$^{-/-}$ mouse models and selective estrogen receptor modulators revealed essential roles of both ERα and ERβ in mouse ectopic lesion development (Burney, 2013; Zhao et al., 2015). However, each ER isoform has a unique expression pattern between endometriotic tissues and normal endometrium. In the case of ERα, it is controversial whether ERα has an endometriotic tissue-specific pattern (Han and O'Malley, 2014). In contrast to ERα, however, the mRNA level of ERβ is significantly higher in endometriotic tissues than in normal uterine endometrium (Bulun et al., 2012). Aberrant ERβ levels in endometriotic tissues have been associated with a distinct epigenomic profile in the ERβ genomic locus: a hypomethylated promoter of the ERβ gene was detected in endometriotic tissues compared with normal endometria and correlates with increased ERβ mRNA levels (Xue et al., 2007).

What is the role of ER isoforms in the pathogenesis of endometriosis? Unfortunately, the detailed molecular mechanism regarding specific contribution of each ER isoform in the endometriosis progression is not clearly elucidated yet. Only partial information is available. For examples, ERα$^{-/-}$ mouse with endometriosis revealed that the *ESR1* gene is required for attachment, inflammation, and proliferation of ectopic lesions (Burns et al., 2012). ERβ directly induces Ras-like
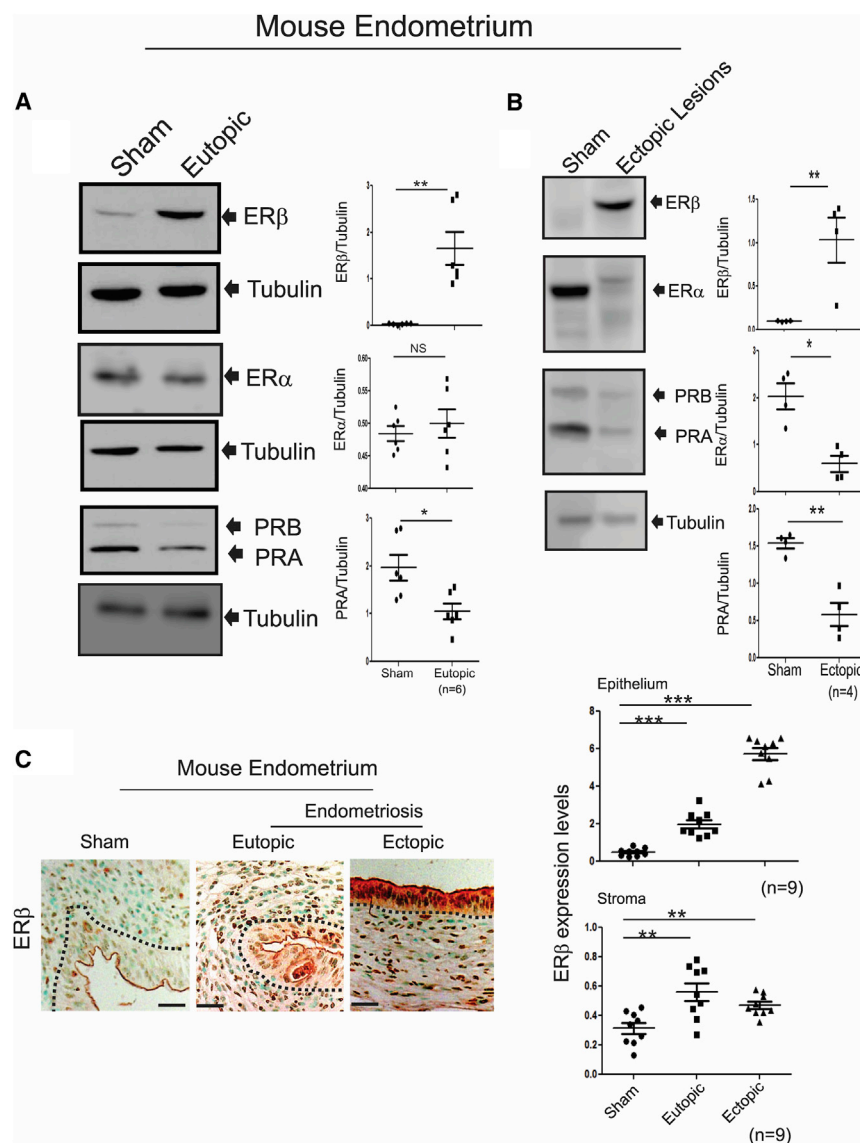
Figure 1. Mouse Endometriotic Tissues Have Elevated Levels of ERβ

(A and B) The expression levels of ERβ, ERα, PR, and tubulin in the uteri of sham-treated C57BL/6J mice and the eutopic endometria (A) and ectopic lesions (B) of C57BL/6J mice with endometriosis.

(C) IHC and quantitative analyses of ERβ levels in the uteri of sham-treated C57BL/6J mice and ectopic and eutopic endometria of C57BL/6J mice with endometriosis.

In all panels, error bars represent ± SD.

observed SRC-1 coactivator isoform, these two drivers of endometriotic disease cooperate to render endometriosis a therapeutically complex disease.

RESULTS

Mouse Endometriotic Tissues Have Elevated ERβ Levels Similar to Those in Human Endometriotic Cells

Human endometriotic cells isolated from endometriosis patients have higher levels of ERβ, but not ERα, than do normal human endometrial cells (Han et al., 2012). Consistent with human endometriotic cells, both eutopic and ectopic endometria from mice with endometriosis also had markedly higher ERβ levels compared with the uteri of sham-treated mice (Figures 1A and 1B). In contrast to ERβ, however, the levels of ERα did not differ in eutopic endometria but were reduced in ectopic lesions compared with sham-treated uteri (Figures 1A and 1B). Levels of PR were reduced in both ectopic lesions and eutopic endometria of mice with endometriosis compared with the uteri of sham-treated mice (Figures 1A

and 1B). Immunohistochemistry (IHC) using an ERβ antibody (validation of its specificity in Figure 4B) revealed elevated ERβ levels in the epithelial and stromal compartments of both ectopic lesions and eutopic endometria compared with those compartments in normal endometrium (Figure 1C). Therefore, the ERβ levels are elevated in endometriotic tissues of mice with endometriosis, similar to the levels observed in human endometriotic cells.

Endometriotic Tissues Have Enhanced ERβ Activity Compared with Normal Endometria

To determine ERβ activity in endometriotic tissues in vivo, we generated an ERβ activity indicator (ERBAI) mouse containing a modified ERβ bacterial artificial chromosome clone that has a Gal4 DNA-binding domain (DBD) instead of its own DBD and a hrGFP reporter controlled by the Gal4-upstream activating sequence (UAS) according to our prior protocol (Han et al., 2009)
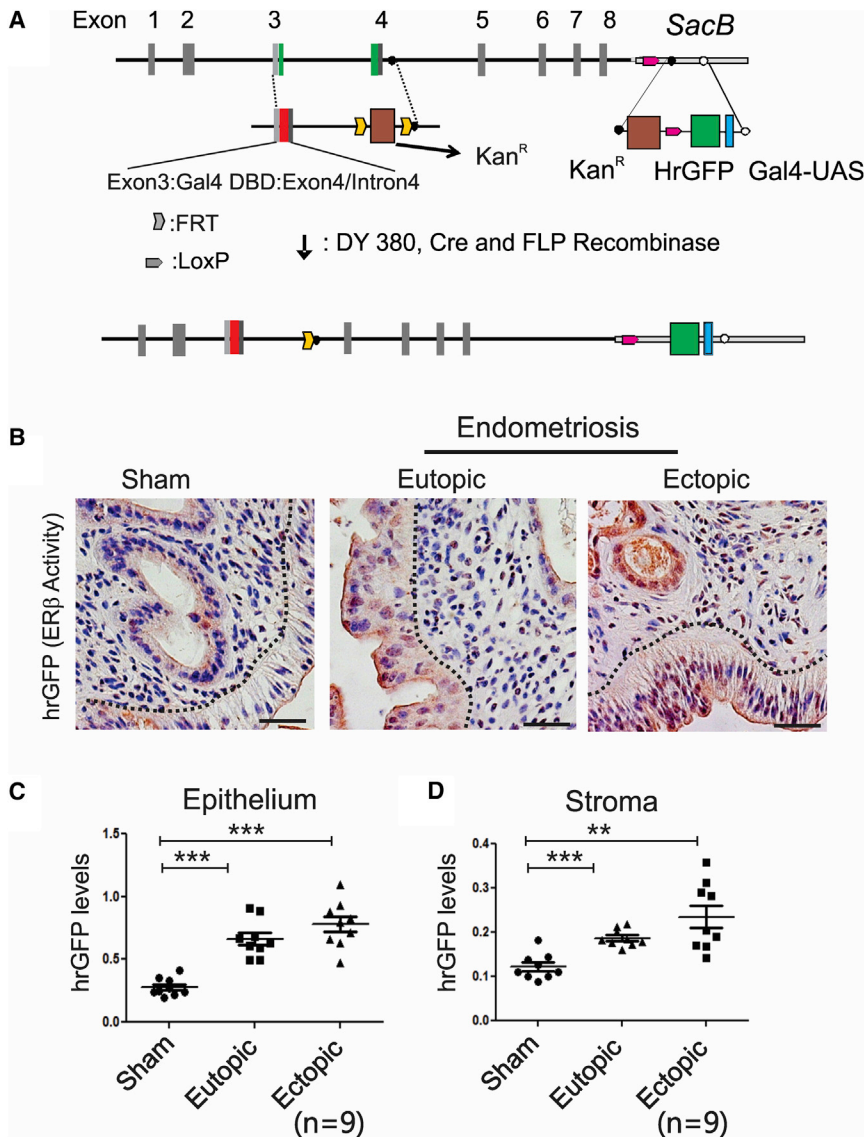
estrogen-regulated growth-inhibitor gene expression in an estrogen-dependent manner to enhance the proliferative activity of endometriotic tissues (Monsivais et al., 2014). In addition, ERβ directly binds to the ERα promoter region to repress ERα gene expression, which can lead to a state of progesterone resistance in the endometriotic tissues by suppressing ERα-mediated progesterone receptor (PR) expression in endometriotic tissues (Bulun et al., 2012). However, we believe the complete repertoire of ERβ functions to be more complicated because greatly elevated levels of ERβ exist in both nuclear and cytoplasmic locations in endometriotic tissues (Cheng et al., 2011). We believed a more detailed investigation should be carried out to fully understand the mechanisms of ERβ action in endometriosis progression.

We propose a cytoplasmic ERB protein network that, in addition to its genomic functions, promotes endometriosis pathogenesis in a non-genomic manner. Together with our previously

Figure 2. Enhanced ERβ Activity Is Detected in the Endometriotic Tissues of Mice with Endometriosis Compared to Normal Endometria

(A) Generation of a modified ERβ bacterial artificial clone that has a Gal4 DNA-binding domain and the Gal4-UAS-hrGFP reporter. DY380, bacterial recombination strain; Kan[R], kanamycin-resistant gene; DBD, DNA-binding domain; Gal4-UAS, Gal4-upstream-activating sequence; FLP, flippase; hrGFP, humanized renilla GFP.

(B) IHC analyses of hrGFP levels in the uteri of sham-treated ERBAI mice and ectopic and eutopic endometria of ERBAI mice with endometriosis.

(C and D) The quantification of hrGFP levels in the epithelial (C) and stromal (D) compartments of each type of endometrium in (B).

In all panels, error bars represent ± SD. See also Figure S1.

Enhanced ERβ Activity Is Required for Ectopic Lesion Growth in Mice with Endometriosis

Although enhanced ERβ activity was detected in endometriotic tissues, it was not clear whether the enhanced ERβ activity was required for ectopic lesion growth. Therefore, PHTPP, an ERβ-selective antagonist (Compton et al., 2004), was employed to address it. Ectopic lesions were surgically developed in ovariectomized C57BL/6J mice containing an Estradiol (E2) pellet. On the 21st day after endometriosis induction, PHTPP or vehicle was administered to endometriosis-induced mice (Figure S2A). Compared with vehicle treatment, PHTPP treatment significantly suppressed ectopic lesion growth in mice with endometriosis (Figure 3A). For the stimulation of ectopic lesion growth, endometriotic tissues recruit immune cells (CD163-positive monocyte/macrophage cells) to enhance immune cell-mediated cytokine signaling (Figure S2B, arrowhead). However, PHTPP-treated ectopic lesions did not recruit immune cells compared with vehicle-treated ectopic lesions (Figure S2B). In addition, PHTPP treatment clearly diminished ERβ activity in the epithelial and stromal compartments of ectopic lesions and the eutopic endometrium of mice with endometriosis compared with vehicle (Figures 3B and 3C). Collectively, PHTPP inhibits ERβ activity, which leads to endometriotic lesion growth in mice with endometriosis.

Anti-apoptosis signaling and the acceleration of proliferation are typical molecular properties associated with the survival of endometriotic tissues (Pellegrini et al., 2012; Salmassi et al., 2011). Because endometriotic tissues consist of epithelial and stromal compartments, signaling communication between these compartments plays an essential role in endometriotic lesion progression (Kim et al., 2013). Therefore, functional defects

(Figures 2A and S1). Therefore, Gal4-ERβ binds to Gal4-UAS, transcribing hrGFP gene expression in response to hormone. Detailed information regarding the generation and validation of the ERBAI mouse model is described in the Supplemental Information (Figure S1).

To investigate potential alterations in ERβ activity in endometriotic tissues during endometriosis progression, endometriosis was surgically induced using an ERBAI mouse model through autotransplantation; to monitor ERβ activity, the hrGFP levels were determined in ectopic and eutopic endometria of ERBAI mice with endometriosis and in the uteri of sham-treated ERBAI mice. Elevated hrGFP levels were detected in epithelial and stromal cells of ectopic and eutopic endometria compared with those found in the normal uterus of sham-treated ERBAI mice (Figures 2B– 2D). Therefore, enhanced ERβ activities were detected in the stromal and epithelial compartments of endometriotic tissues compared with normal endometria.

(legend on next page)

involving hyperproliferation and anti-apoptosis signaling in either compartment in endometriotic tissues should impair cellular processes in the other compartment, ultimately leading to the suppression of ectopic lesion growth. PHTPP treatment, compared with vehicle treatment, reduced proliferative activity as determined by Ki-67 in the epithelial, but not stromal, compartments of ectopic lesions in C57BL/6J mice with endometriosis (Figure 3D). In the case of apoptosis signaling as determined by cleaved caspase 8 levels, PHTPP treatment significantly enhanced apoptotic signaling in both the epithelial and stromal compartments of ectopic lesions of C57BL/6J mice with endometriosis compared with vehicle treatment (Figure 3E). In addition to ectopic lesions, PHTPP suppressed proliferation and anti-apoptosis signaling in the epithelium and inhibited proliferation in stromal cells in the eutopic endometria of mice with endometriosis (Figures 3F and 3G).

In addition to ERβ, PHTPP can inhibit ERα activity in vivo, although its effects on ERα are minimal (Compton et al., 2004). To address this issue, the levels of mouse uterine ERα direct target genes (such as PR, CDKN1A, and ERRFI1) were examined in ovariectomized mice upon E2 and/or PHTPP treatment. PHTPP partly reduced the expression of direct ERα target genes stimulated by E2 (Figures S2C–S2E). Interestingly, a female mouse fertility assay revealed that PHTPP did not reduce reproductive activity in female mice, whereas an ERα-selective antagonist, MPP dihydrochloride, significantly reduced the fertility of female mice compared with vehicle treatment (Figure S2F). Therefore, PHTPP does not disrupt the fertility of female mice, though it partly suppresses uterine ERα activity. In contrast to the effects of PHTPP, ERB-041 (ERβ-specific agonist) treatment enhanced the mouse ectopic lesion growth compared with vehicle (Figure S2G).

To address the effects of ERB-041 and PHTPP in human endometriotic lesion growth, we employed two types of human endometrial cells: EMosis-CC/TERT cells, which are immortalized human endometriotic epithelial cells isolated from ovarian endometriomas (Bono et al., 2012), and immortalized human endometrial stromal cells (iHESCs) (Krikun et al., 2004). For simplification, EMosis-CC/TERT cells are called immortalized human endometriotic epithelial cells (iHEECs) hereafter. For non-invasive bioluminescence imaging analyses of ectopic lesions in SCID mice, luciferase reporters expressing iHEECs (iHEECs/Luc) and iHESCs (iHESCs/Luc) were generated using a lentiviral expression system. To induce endometriosis, a mixture of epithelial and stromal cells (iHEECs/Luc plus iHESCs/Luc) was injected into ovariectomized SCID mice with an E2 pellet. On the 21st day after endometriosis induction, endometriosis-induced SCID mice were treated with ERB-041 or PHTPP for

another 21 days (Figure S2A). Ectopic lesion image analyses revealed that ERB-041 treatment stimulated human ectopic lesion growth, whereas PHTPP treatment decreased ectopic lesion growth in SCID mice (Figure S2H). Moreover, ERBAI mice with endometriosis also revealed that ERB-041 enhanced ERβ activity in ectopic lesions compared with vehicle treatment (Figure S2I). Collectively, enhanced ERβ activity is required for the pathogenesis of endometriosis (Table 1).

Loss of ERβ Function Suppresses Ectopic Lesion Growth in Mice with Endometriosis

To directly investigate the loss of ERβ function in the pathogenesis of endometriosis, endometriosis was surgically induced via the auto-transplantation of uterine tissue using ERβ$^{-/-}$ (Krege et al., 1998) and wild-type (WT) mice. The sizes of the ERβ$^{-/-}$ ectopic lesions were reduced significantly compared with WT ectopic lesions (Figure 4A). IHC using an ERβ antibody (Saji et al., 2000) validated the fact that ERβ$^{-/-}$ ectopic lesions did not exhibit ERβ expression compared with WT ectopic lesions (Figure 4B).

To investigate how loss of the ERβ function impacts ectopic lesion progression, cell proliferation and apoptotic signals in each type of ectopic lesion were examined. The reduced levels of epithelial, but not stromal, proliferation were detected in ERβ$^{-/-}$ ectopic lesions compared with WT ectopic lesions (Figure 4C and Table 1). In contrast to proliferation, however, loss of ERβ functions significantly elevated epithelial, but not stromal, apoptosis in ERβ$^{-/-}$ ectopic lesions (Figure 4D and Table 1).

Regarding eutopic endometrium, loss of ERβ function did not impair the proliferation of ERβ$^{-/-}$ eutopic endometria compared with WT eutopic endometria during endometriosis progression (Figure 4E and Table 1). However, apoptosis signaling was elevated in both compartments of the eutopic endometrium in the absence of the ERβ gene compared with WT eutopic endometrium (Figure 4F and Table 1).

Gain of ERβ Function Stimulates Ectopic Lesion Growth in Mice with Endometriosis

To mimic ERβ elevation in human and mouse endometriotic tissues, an endometrium-specific ERβ-overexpressing mouse model was generated and validated (Figure S3). For simplification, ROSA$^{LSL:ERβ/+}$ monogenic mice, which do not express exogenous ERβ, and endometrium-specific ERβ-overexpressing (ROSA$^{LSL:ERβ/+}$:PR$^{CRE/+}$) bigenic mice are hereafter referred to as control and ERβ:OE mice, respectively (Figure S3A).

To determine whether endometrium-specific ERβ overexpression impacts ectopic lesion growth, endometriosis was surgically induced by auto-transplantation using ovariectomized

Figure 3. ERβ-Specific Antagonist Regresses Ectopic Lesion Growth
(A) Ectopic lesions isolated from C57BL/6J mice with endometriosis subcutaneously treated with vehicle or PHTPP.
(B and C) IHC and quantitative analyses of hrGFP levels in ectopic lesions (B) and eutopic endometria (C) of ERBAI mice with endometriosis subcutaneously treated with vehicle or PHTPP.
(D and E) IHC and quantitative analyses of the expression patterns of Ki-67 (D) and cleaved CSP8 (E) in ectopic lesions of C57BL/6J mice with endometriosis subcutaneously treated with vehicle or PHTPP.
(F and G) IHC and quantitative analyses of the levels of Ki-67 (F) and cleaved CSP8 (G) in the eutopic endometria of C57BL/6J mice with endometriosis subcutaneously treated with vehicle or PHTPP. PLC, percentage of labeled cells; CSP8, caspase 8.
In all panels, error bars represent ± SD. See also Figure S2.

Table 1. Proliferation and Apoptosis in Ectopic Lesions and Eutopic Endometria of PHTPP-Treated, ERβ$^{-/-}$, and ERβ:OE Mice with Endometriosis

Cellular Process	Type of Endometrium	Compartment	PHTPP	ERB$^{-/-}$	ERB:OE
Proliferative activity	ectopic lesions	epithelium	−	−	+
		stromal	0	0	+
	eutopic endometrium	epithelium	−	0	+
		stromal	−	0	+
Apoptosis signaling	ectopic lesions	epithelium	+	+	−
		stromal	+	0	−
	eutopic endometrium	epithelium	+	+	0
		Stromal	0	+	0
Ectopic lesion volume			−	−	+

+: increased, 0: no change, −: decreased compared to vehicle treatment, WT, or control mice.

control and ERβ:OE mice containing E2 pellets. ERβ:OE ectopic lesions had much larger volumes than did control ectopic lesions in mice with endometriosis (Figure 5A). Exogenous Flag/Myc-tagged ERβ expression and elevated levels of total ERβ were determined in ERβ:OE ectopic lesions compared with control ectopic lesions (Figures 5B and S3). Thus, elevated ERβ levels in ectopic lesions enhanced ectopic lesion growth. The overexpression of nuclear receptors can induce ligand-independent effects (Weigel and Zhang, 1998). However, neither control nor ERβ:OE endometrial tissue fragments successfully developed into ectopic lesions in ovariectomized mice without the administration of an E2 pellet (Figure S3F). Therefore, gain-of-ERβ-function-mediated stimulation of ectopic lesion growth is an estrogen-dependent process.

The proliferative activity was significantly elevated in both the epithelial and stromal compartments of ERβ:OE ectopic lesions compared with control ectopic lesions (Figure 5C and Table 1). However, apoptosis was significantly reduced in both the epithelial and stromal compartments of ERβ:OE ectopic lesions compared with control lesions (Figure 5D and Table 1).

In addition to ectopic lesions, both the epithelial and stromal compartments of the ERβ:OE eutopic endometrium demonstrated enhanced proliferative activity compared with control eutopic endometrium (Figure 5E and Table 1). However, no alteration in apoptosis signaling was detected in ERβ:OE eutopic endometrium compared with control eutopic endometrium (Figure 5F and Table 1). Thus, the eutopic endometria of endometriosis patients appear primarily to be in a hyperproliferative state due to elevated ERβ levels. Notably, breeding trials designed to assess mating success revealed that ERβ:OE mice were infertile compared with control mice (Figure S3G). Moreover, ERβ-overexpressing iHESCs lose their decidualization response because, as compared with parental iHESCs, the induction of decidual cell marker genes, such as insulin-like growth factor-binding protein 1 and prolactin, was significantly reduced upon estradiol-medroxyprogesterone-cAMP treatment (Figures 5G–5I). Therefore, endometriosis-associated ERβ overexpression in eutopic endometrium might impair the decidualization process in women with endometriosis, leading to infertility.

ERβ Interacts with the Cytoplasmic Apoptosis and Inflammasome Machinery in Ectopic Lesions to Enhance Ectopic Lesion Survival

To further dissect the molecular mechanisms of ERβ in endometriosis progression, Flag-tagged ERβ-containing protein complexes were immunoprecipitated (IPed) from the eutopic endometria of ERβ:OE mice with endometriosis using a Flag antibody. In IP/Mass analyses, a primary consideration is to separate out proteins that non-specifically interact with beads from the list of proteins that are associated with the target protein. For this purpose, we employed control mice that had the same genetic background as ERβ:OE mice and had extremely low levels of Flag-tagged exogenous ERβ compared with ERβ:OE mice (Figure 6A). Therefore, proteins co-IPed with the Flag antibody from endometriotic tissues of control mice are considered as non-specific bead-binding proteins. To specifically identify ERβ-interacting proteins, proteins IPed from the eutopic endometria of control mice were removed from the proteins that IPed from the eutopic endometria of ERβ:OE mice. Gene Ontology analyses with endometriotic tissue-specific ERβ-interacting proteins revealed that large numbers of proteins involved in inflammation and apoptosis signaling were specifically co-IPed with ERβ from ERβ:OE eutopic endometrium (Figures S4A and S4B). To validate these interactions in ectopic lesions, the ERβ complex was isolated from control and ERβ:OE ectopic lesions using a Flag antibody, and ERβ-interacting proteins were analyzed further by western blot analyses (Figure 6A). Western blot analyses revealed that only a very weak Flag-ERβ signal was detected in control ectopic lesions that developed in ROSA$^{LSL:ERβ/+}$ monogenic mice; this small amount is likely attributable to the leaky expression of exogenous ERβ in the ROSA$^{LSL:ERβ/+}$ monogenic mouse (Figure 6A).

Apoptosis signal-regulating kinase 1 (ASK-1) was found to interact prominently with ERβ in ectopic lesions (Figure 6A). ASK-1 is a component of tumor necrosis factor alpha (TNF-α)-induced apoptosis complex I, and its activation is required for TNF-α-induced apoptosis in multiple cell types (Tobiume et al., 2001). In addition to ASK-1, serine/threonine kinase receptor-associated protein (STRAP) and 14-3-3 were also specifically co-IPed with ERβ from the ERβ:OE ectopic lesions, but not from control ectopic lesions (Figure 6A). To prevent TNF-α-induced apoptosis, STRAP and 14-3-3 proteins interact with ASK-1 to disrupt associations between TNF receptor-associated factor 2 (TRAF2) and ASK-1 upon TNF-α stimulation (Hatai et al., 2000). These data imply that ERβ may induce ASK-1/STRAP/14-3-3 complex formation to prevent the activation of TNF-α/ASK-1-mediated apoptosis in endometriotic tissues. To validate this hypothesis, the levels of ASK-1 phosphorylation at Thr845 were determined for each type of ectopic lesion because ASK-1 phosphorylation at Thr845 is associated with ASK-1 activation to enhance TNF-α-induced apoptosis (Tobiume et al., 2002). The phospho-Thr845 ASK-1 levels were significantly reduced in ERβ:OE ectopic lesions compared with control ectopic lesions without alternation of total ASK-1 levels (Figures 6B–6D). In contrast, the levels of total ASK-1 and phospho-ASK-1 were significantly elevated in ERβ$^{-/-}$ ectopic lesions compared with WT ectopic lesions (Figures S5A– S5C). Collectively, ERβ induced ASK-1/STRAP/14-3-3 complex formation
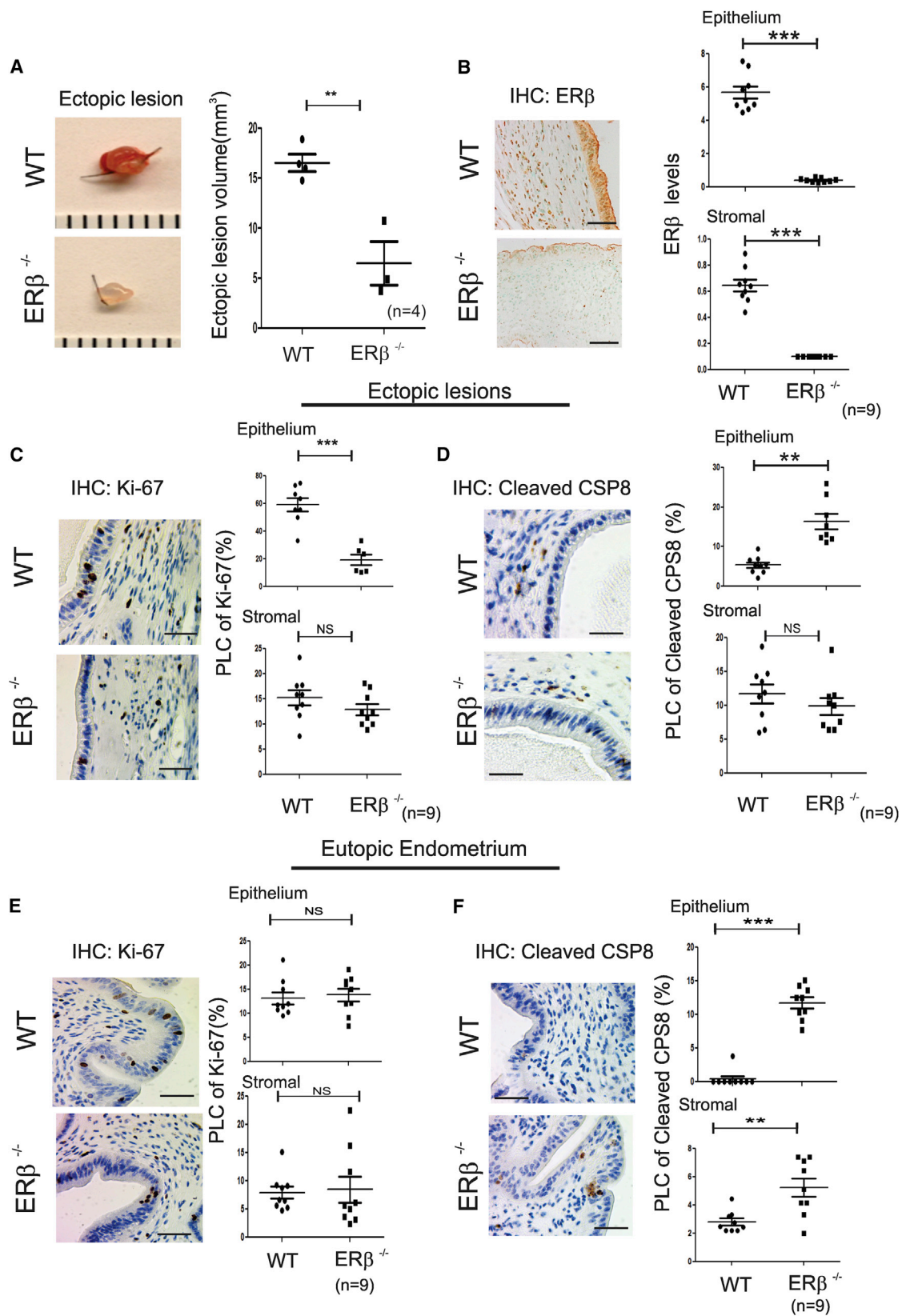
(legend on next page)

to prevent ASK-1 activation in ectopic lesions, thereby promoting ectopic lesion survival. TNF-α-induced ASK-1 activation also increases mitochondrial cytochrome *c* levels to activate caspase 9 (Hatai et al., 2000). The cytochrome *c* levels in ERβ:OE ectopic lesions were significantly lower than those in control ectopic lesions (Figure 6E). Therefore, the gain of ERβ function prevented the TNF-α/ASK-1/cytochrome *c* signaling pathway in ectopic lesions to promote lesion survival.

After the initiation of TNF-α-induced apoptosis by apoptosis complex I, the tumor necrosis factor receptor (TNFR) 1-associated death domain (TRADD) protein, a component of complex I, is shuttled from TNFR to the cytoplasm and then interacts with the Fas-associated via death domain (FADD) protein and caspase 8 to generate apoptosis complex II to amplify TNF-α-induced apoptosis (Micheau and Tschopp, 2003). The endometriotic 70 kDa SRC-1 isoform also interacts with caspase 8 to inhibit caspase 8 activation in ectopic lesions to promote their survival (Han et al., 2012). Interestingly, we also found that ERβ interacted with caspase 8 and this SRC-1 isoform in ectopic lesions (Figures 6A and S4C). Moreover, ERβ:OE ectopic lesions contained significantly reduced levels of cleaved caspase 8 compared with control ectopic lesions (Figure 5D). Therefore, we suggest that ERβ also interacts with caspase 8 along with the SRC-1 isoform; this combined interaction strongly inhibits caspase 8 activation in ectopic lesions to effectively prevent activation of TNF-α-induced apoptosis complex II in ectopic lesions. However, this SRC-1 isoform/ERβ/caspase 8 complex did not interact with components of TNF-α-induced apoptosis complex I and the apoptosome in ectopic lesions (Figure S4D).

To validate synergism between the SRC-1 isoform and ERβ for the progression of endometriosis, a combination of Gossypol that reduces the transcriptional activity and stability of SRC-1 (Wang et al., 2011) and PHTPP was employed to suppress ectopic lesion growth in mice with endometriosis. This combination of Gossypol and PHTPP treatment significantly reduced ectopic lesion growth compared with individual treatments (Figures 6F and 6G). Therefore, cooperative interactions between the ERβ and SRC-1 isoforms effectively appear to drive the pathogenesis of endometriosis.

The cytochrome *c* effectively induces the formation of the apoptosome, which consists of caspase 9 and apoptotic peptidase-activating factor 1 (APAF-1), to activate caspase 9 (Bratton and Salvesen, 2010). In ERβ:OE ectopic lesions, the interaction of caspase 9 and APAF-1 was not detected (Figures 6H and S5D). The cleaved caspase 9 levels were significantly reduced in ERβ:OE ectopic lesions compared with those in control ectopic lesions (Figure 6I). In ERβ$^{-/-}$ ectopic lesions, however, the interaction of caspase 9 and APAF-1 was detected (Figure S5D). These data suggest that ERβ prevented TNF-α-induced apoptosome formation in endometriotic cells by disrupting the interaction of caspase 9 and APAF-1 through a competitive ERβ interaction with caspase 9. Collectively, ERβ synergistically inhibited the activation of apoptosis complex I and complex II and apoptosome formation in ectopic lesions to effectively prevent TNF-α-induced apoptosis in endometriotic tissues for ectopic lesions survival.

Caspase 1 and the NLR family pyrin domain-containing 3 (NALP3) were also co-IPed with ERβ from ERβ:OE ectopic lesions (Figures 6H and S4B). Both caspase 1 and NALP3 are components of the inflammasome, which is involved in the maturation of IL-1β formation from pro-IL-1β (Willingham et al., 2009). Interestingly, the NALP3-mediated inflammasome has an essential role in endometriosis progression because NALP3$^{-/-}$ ectopic lesion volume was significantly reduced compared with WT ectopic lesions of mice with endometriosis (Figure 6J). IL-1β is a key cytokine involved in both the adhesion and proliferation of endometrial cells (Cao et al., 2005; Sillem et al., 1999). ERβ:OE ectopic lesions had higher IL-1β levels than control ectopic lesions because cleaved caspase 1 levels were highly elevated in ERβ:OE ectopic lesions compared with controls (Figures 6K and 6L). However, cleaved caspase 1 and IL-1β levels were reduced in ERβ$^{-/-}$ ectopic lesions compared with WT lesions of mice with endometriosis (Figure S5A). Therefore, the combinational interactions of ERβ with caspase 1 and NALP3, the activation of caspase 1, and the elevation of IL-1β levels in ectopic lesions supported our conclusion that ERβ also enhances inflammasome activity in ectopic lesions for their survival.

This Flag-ERβ-interacting protein network may not accurately recapitulate the endogenous ERβ-interacting protein network in endometriotic tissues because this network was generated by overexpressed exogenous ERβ. To further address this issue, ERβ complexes were isolated from ectopic lesions of C57BL/6J mice with endometriosis because these ectopic lesions had elevated endogenous ERβ levels (Figure 1B). The SRC-1 isoform, caspase 8, caspase 9, caspase 1, and ASK-1 were also co-IPed with endogenous ERβ from ectopic lesions similar to exogenous Flag-ERβ, though the IP efficiency of ERβ antibody (SC-8794, Santa Cruz) is low (Figure S4E). Therefore, we concluded that the overexpressed ERβ complex is similar to the endogenous ERβ complex in endometriotic tissues.

Gain of ERβ Function Prevents TNF-α-Induced Apoptosis and Enhances Proliferation, Invasion, and Adhesion Activities of Immortalized Human Endometriotic Epithelial Cells

To investigate the functions of ERβ and the ERβ/SRC-1 isoform complex, iHEECs stably expressing ERβ (iHEECs/ERβ) and the SRC-1 isoform (iHEECs/SRC-1Iso) were generated separately and together (iHEECs/ERβ/SRC-1Iso) (Figure 7A, in bottom). TNF-α treatment increased the levels of cleaved caspase 8 and cleaved caspase 3 in control iHEECs compared with vehicle

---

Figure 4. The Loss of ERβ Function Prevents Ectopic Lesion Growth

(A) Ectopic lesions isolated from C57BL/6J (WT) and ERβ$^{-/-}$ mice with endometriosis.

(B) IHC analyses and quantification of the ERβ levels in ectopic lesions isolated from WT and ERβ$^{-/-}$ mice with endometriosis.

(C–F) IHC and quantitative analyses of Ki-67 (C and E) and cleaved CSP8 (D and F) in the epithelial and stromal compartments of ectopic lesions (C and D) and eutopic endometria (E and F) of WT and ERβ$^{-/-}$ mice with endometriosis.
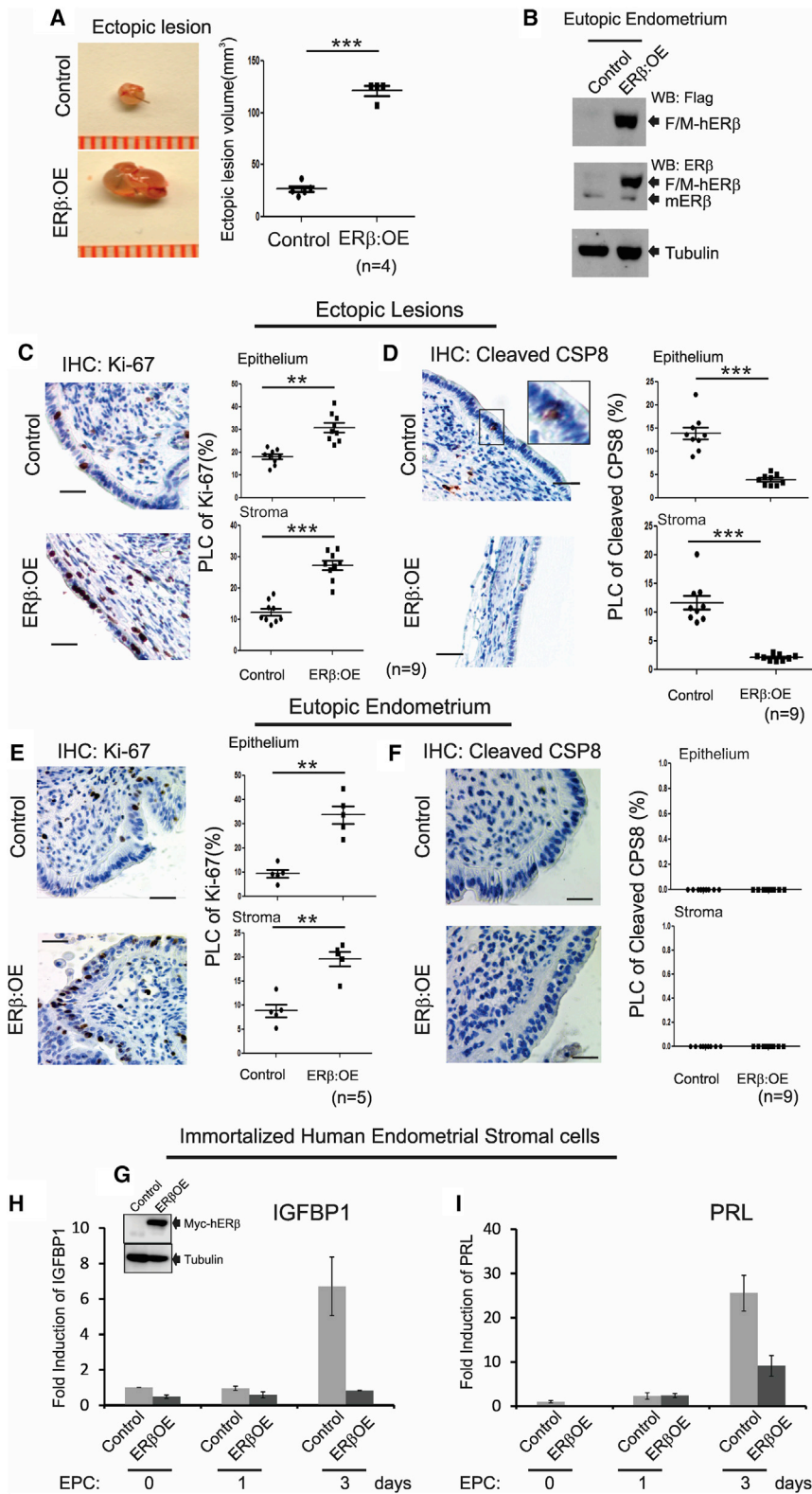
In all panels, error bars represent ± SD.

Figure 5. The Gain of ERβ Function Stimulates Ectopic Lesion Growth

(A) Ectopic lesions isolated from control and ERβ:OE mice with endometriosis.

(B) Exogenous Flag/Myc-tagged human ERβ (F/M-hERβ) protein levels in the eutopic endometria of control and ERβ:OE mice with endometriosis. mERβ, endogenous mouse ERβ.

(C–F) IHC and quantitative analyses of Ki-67 (C and E) and cleaved CSP8 (D and F) in the epithelial and stromal compartments of ectopic lesions (C and D) and eutopic endometria (E and F) of control and ERβ:OE mice with endometriosis. Higher magnification views of the boxed regions are shown.

(G) Exogenous Myc-tagged human ERβ (Myc-hERβ) protein levels in iHESCs/ERβ as determined with a Myc antibody.

(H and I) The quantification of relative changes in the mRNA levels of decidualization marker genes, IGFBP1 (H) and PRL (I), in iHESCs (Control) and iHESCs/ERβ (ERβOE) upon estrogen/medroxyprogesterone/db-cAMP (ECP) treatment on the indicated day.

In all panels, error bars represent ± SD. See also Figures S3.
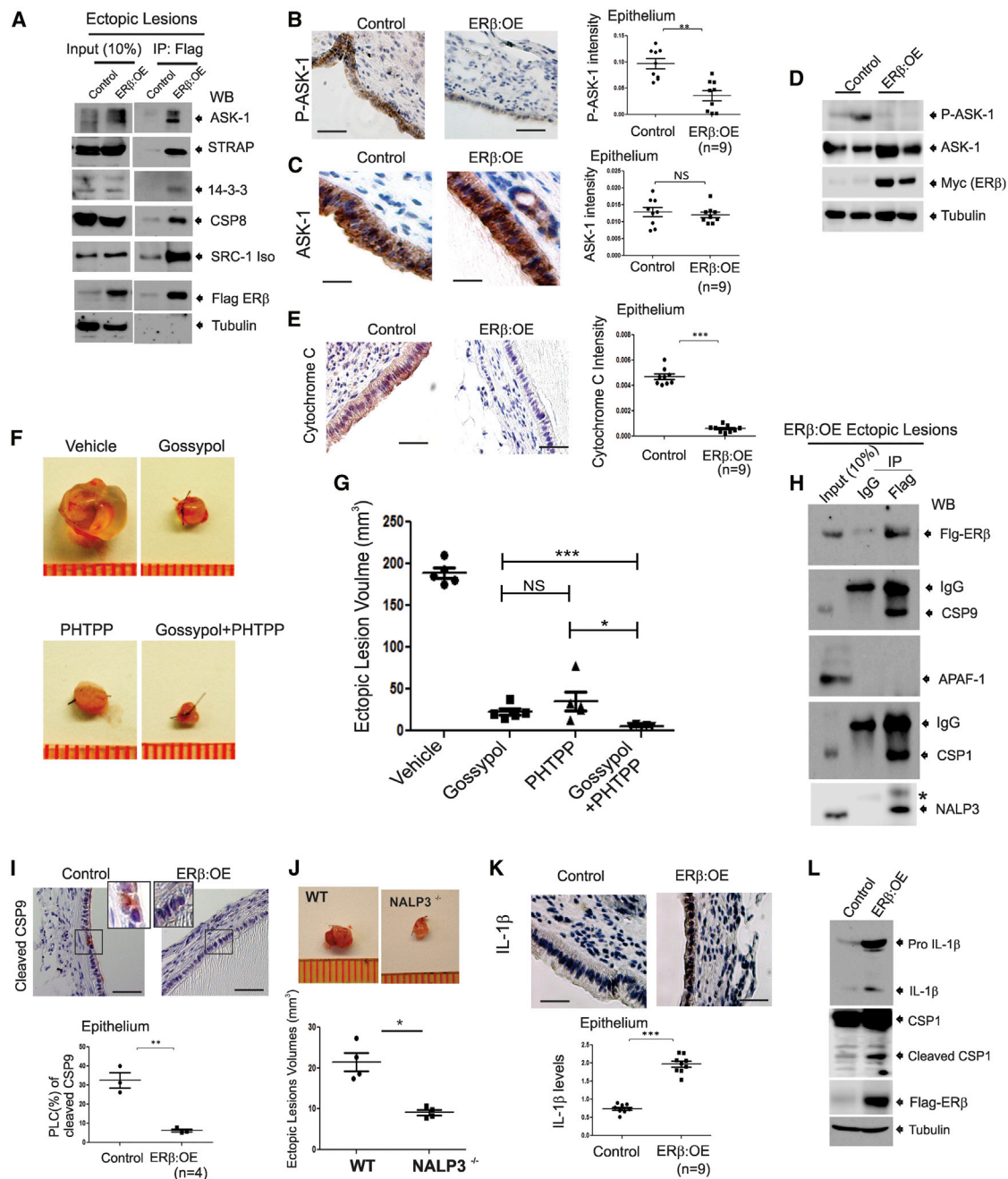
Figure 6. ERβ Interacts with TNF-α-Induced Apoptosis Complexes and the Inflammasome in Endometriotic Tissues of Mice with Endometriosis

(A) Flag-ERβ complexes IPed with a Flag antibody from ectopic lesions of control and ERβ:OE mice with endometriosis followed by western blotting (WB) with antibodies against ASK-1, STRAP, 14-3-3, CSP8, SRC-1, Flag, and tubulin.

(B and C) IHC and quantitative analyses of phospho-Thr845-ASK-1 (P-ASK-1) (B) and total ASK-1 (C) in control and ERβ:OE ectopic lesions.

(D) Western blot analyses of phospho-Thr845-ASK-1 (P-ASK-1), total ASK-1, ERβ, and tubulin in control and ERβ:OE ectopic lesions.

(E) IHC and quantitative analyses of cytochrome *c* levels in control and ERβ:OE ectopic lesions.

(F and G) Regression of ectopic lesion growth in endometriosis-induced C57BL/6J mice subcutaneously treated with Gossypol, PHTPP, or their combination compared to vehicle (F). Quantification of ectopic lesion volume in (F) is shown in the graph (G).

(H) The IPed Flag-ERβ complex from ERβ:OE ectopic lesions with a Flag antibody or IgG followed by western blotting with antibodies against Flag, CSP 9, APAF-1, CSP1, and NALP3. *, non-specific protein.

(I) IHC and quantitative analyses of cleaved CSP9 levels in control and ERβ:OE ectopic lesions. Higher-magnification views of the boxed regions are shown.

(J) Ectopic lesions isolated from C57BL/6J (WT) and NALP3[−/−] mice with endometriosis.

(legend continued on next page)

(Figure 7A). However, TNF-α treatment did not enhance the levels of the above apoptosis markers in iHEECs/ERβ, in iHEECs/SRC-1Iso, or in the combined iHEECs/SRC-1Iso/ERβ (Figure 7A). Therefore, ERβ and ERβ/SRC-1 isoform complex effectively prevented TNF-α-induced apoptosis. The gain of ERβ, but not the SRC-1 isoform, function elevated the IL-1β levels in iHEECs in the presence or absence of TNF-α treatment (Figure 7A). These data were also generated via the artificial overexpression of ERβ. To support that gain of ERβ function is not artificial, primary human endometriotic stromal cells isolated from human endometriosis patients were employed because these cells also had elevated levels of endogenous ERβ compared with normal human endometrial stromal cells (Figure S4F). Primary human endometriotic stromal cells also have elevated levels of IL-1B and anti-apoptosis signaling upon TNF-α treatment compared with normal endometrial stromal cells (Figure S4G). Therefore, ERβ plays a critical role in anti-apoptosis signaling and inflammasome activation in ectopic lesions. In addition, ERβ enhanced the cell adhesion and proliferative activities of iHEECs compared with control cells in the presence of TNF-α (Figures 7A and 7B). Therefore, it is likely that the increased IL-1β observed with elevated ERβ induces adhesion and proliferation activities of endometrial tissue fragments in the peritoneal area of endometriosis patients to initiate ectopic lesion development. In addition to IL-1β, the levels of several cytokines (such as MIP-2, IL-16, MIP-1a, MCP-5, TREM-1, and BLC) were significantly elevated in ERB:OE ectopic lesions compared with control ectopic lesions (Figures S6A and S6C). Previous studies also revealed that levels of these cytokines are elevated in the peritoneal fluid of women with endometriosis (Ahn et al., 2015). In contrast, some cytokine levels (MIG, M-CSF, TNF-α, KC, and IP-10) were reduced in ERB:OE ectopic lesions compared to control ectopic lesions (Figures S6A and S6B). Collectively, the gain of ERβ function broadly alters the cytokine milieu in ectopic lesions in concert with promotion of endometriotic lesion growth. Consistent with SRC-1 isoform, ERβ also increased the expression of EMT markers, such as Slug and Snail, and enhanced invasion activity in iHEECs (Figures 7A and 7C). Therefore, the increased EMT and invasion activity of ectopic lesions again occurs through the effective cooperation of the ERβ and SRC-1 isoforms. However, vascular endothelial growth factor (VEGF) levels were not changed in ERβ:OE ectopic lesions and iHEECs/ERβ compared with their controls (Figures S6D and S6E). Therefore, the angiogenesis of ectopic lesions is not regulated by ERβ.

To further support the gain of ERβ function in human ectopic lesion development, iHEECs/Luc, iHEECs/ERβ/Luc, and iHESCs/Luc were employed for non-invasive bioluminescence imaging analysis of ectopic lesion growth in SCID mice. To induce endometriosis, a mixture of iHESCs/Luc plus iHEECs/ERβ/Luc was injected into ovariectomized SCID mice with an E2 pellet. For controls, a mixture of iHESCs/Luc and iHEECs/Luc was injected. Bioluminescence image analysis on

injection day 0 revealed that similar amounts of human endometrial cells for each group were injected into recipient SCID mice (Figure 7D). Comparative bioluminescent analysis on the 21st day after injection revealed that human ectopic lesions with ERβ overexpression exhibited stronger bioluminescent activity compared with control ectopic lesions (Figure 7E). Therefore, ERβ enhanced the in vivo survival rate of human endometriotic cells and promoted their development into human ectopic lesions in SCID mice.

In addition to ERβ, ERα plays an essential role in the pathogenesis of endometriosis in the mouse model (Burns et al., 2012). To determine the functional difference between ERα and ERβ in endometriosis progression, iHEECs expressing Myc-tagged human ERα genes (iHEECs/ERα) were generated (Figure S7A). In contrast to iHEECs/ERβ, however, gain of ERα function did not prevent TNF-α-induced apoptosis signaling and did not induce IL-1β expression, proliferative activity, or expression of EMT markers (Slug and Snail) and VEGF in iHEECs/ERα compared with parental iHEECs upon TNF-α treatment (Figure S7A).

Unlike ERβ, therefore, ERα is not involved directly in the evasion of immune surveillance or in the invasion and IL-1β-mediated proliferation of ectopic lesions. For the combination of ERα and ERβ, ERα did not interfere with ERβ-mediated anti-apoptotic activity in iHEECs upon TNF-α treatment (Figure S7B). In fact, ERα inhibited ERβ-mediated IL-1β production (Figure S7B). Therefore, ERα might be involved in the negative regulation of ERβ-mediated inflammasome activation.

Taken together, the gain of ERβ and SRC-1 isoform function in endometrial fragments generated by retrograde menstruation prevents TNF-α-induced apoptosis complex activity to evade immune surveillance. After evasion, ERβ interacts with the inflammasome complex to induce IL-1β in endometrial fragments that have evaded immune surveillance to facilitate attachment to and growth at target sites (Figure 7F). In addition, ERβ also induces EMT and invasion activity in cooperation with the SRC-1 isoform to establish endometriotic lesions (Figure 7F).

DISCUSSION

ERβ Has Non-genomic Action for Anti-apoptosis and Inflammasome Activation

The physiological effects of estrogen are mediated by estradiol binding to one of the ER isoforms, ERα and ERβ. Estrogen-liganded ER isoform then binds to specific DNA sequences called estrogen response elements. Interestingly, phenotype analyses of ERα$^{-/-}$, ERβ$^{-/-}$, ERα$^{-/-}$:ERβ$^{-/-}$ bigenic mouse models have revealed that ER isoforms have overlapping but also unique roles in estrogen-dependent action in vivo (Walker and Korach, 2004). For their unique function, ERα and ERβ have different transcriptional activities in certain ligand, cell-type, and promoter contexts. In the case of endometriotic tissues, both ER isoforms are expressed in endometriotic tissues and required for endometriotic lesion growth. The gain of ERβ

(K) IHC and quantitative analyses of IL-1β levels in control and ERβ:OE ectopic lesions.

(L) Western blot analyses of levels of IL-1β, CSP1, Flag-tagged ERβ, and tubulin (as a protein loading control) in ectopic lesions of control and ERβ:OE mice with surgically induced endometriosis.

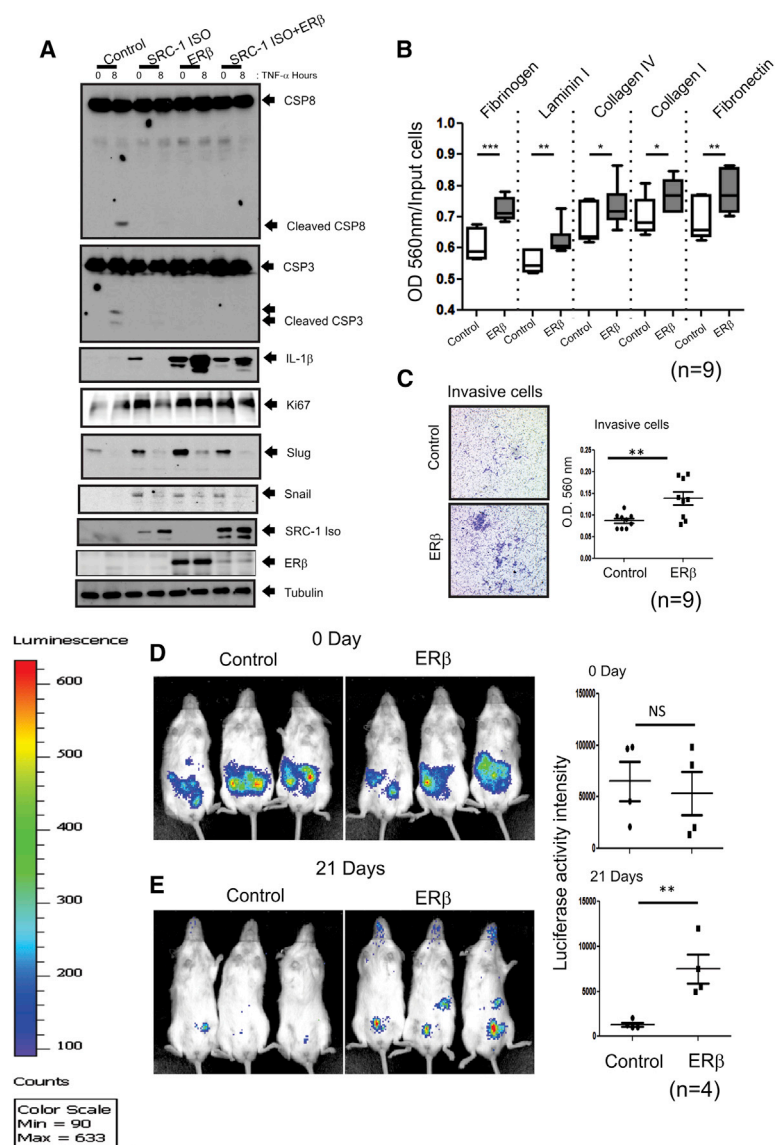In all panels, error bars represent ± SD. See also Figures S4 and S5.

Figure 7. Gain of ERβ Function Prevents TNF-α-Induced Apoptosis Signaling but Stimulates Proliferation, Adhesion, and Invasion Activities of Human Endometriotic Cells

(A) Levels of cleaved CSP8, cleaved CSP3, IL-1β, Ki67, Slug, Snail, and SRC-1 isoform (determined by a Flag antibody); ERβ (determined using a Myc antibody); and tubulin in iHEECs (Control), iHEECs/SRC-1Iso (SRC-1ISO), iHEECs/ERβ (ERβ), or iHEECs/SRC-1Iso/ERβ (SRC-1ISO+ERβ) upon 50 ng/ml TNF-α plus 10 μg/ml cycloheximide treatment for 0 and 8 hr.

(B) Cell-adhesion activities of paternal iHEECs (Control) and iHEECs/ERβ (ERβ) against various extracellular matrices in the presence of 50 ng/ml TNF-α.

(C) Invasion activities of iHEECs (Control) and iHEECs/ERβ (ERβ) for 2 days using a Transwell plate assay. The amounts of invasive cells in each group were determined using a crystal violet staining protocol and are shown in the graph.

(D and E) Bioluminescence and quantitative analyses of iHEECs/Luc (Control) and iHEECs/ERβ/Luc (ERβ) in SCID mice at 0 (D) and 21 (E) days after the induction of endometriosis.

(F) Working model for the non-genomic action of ERβ in endometriosis progression.

In all panels, error bars represent ± SD. See also Figures S6 and S7.

function study revealed that ERβ prevents apoptosis singling and enhances adhesion, invasion, proliferation, inflammasome activity, and inflammation signaling in ectopic lesions for their growth. The study with ERα$^{(-/-)}$ mice with endometriosis revealed that ERα drives proliferation, adhesion, and angiogenesis and also modulates inflammation signaling in ectopic lesions (Burns et al., 2012). To establish endometriotic lesions, collectively, both ERα and ERβ might synergistically contribute the regulation of proliferation, adhesion, and inflammation signaling in ectopic lesions. However, ERα mainly drives angiogenesis, and ERβ has a predominant role in anti-apoptosis and activation of inflammasome and invasion in ectopic lesions for their survival.

Based on a retrograde menstruation model for endometriosis (Hughesdon, 1958), endometrial tissue and erythrocytes are shed through the fallopian tubes into the peritoneal cavity during menses. In healthy women, refluxed endometrial fragments that appear during retrograde menstruation are cleared by inflammatory mediated cell-death signaling, such as caspase-1-mediated pyroptosis (Miao et al., 2011). However, endometriosis patients have an immunity that prevents them from clearing the refluxed endometrial fragments and then potentiates the development and severity of endometriosis. For survival, endometrial tissue fragments must evade the immune surveillance system, particularly peritoneal macrophages (Nasu et al., 2009). During the early steps of evasion from the immune surveillance system, ERβ generates a cytoplasmic protein network to rapidly prevent TNF-β-mediated apoptosis by inactivating TNF-α-induced apoptosis complex I and II and the apoptosome. We believe that a key synergism exists between ERβ and the SRC-1 isoform during this evasion of the immune surveillance system because the SRC-1 isoform also prevents TNF-α-induced apoptosis in ectopic lesions. Our combined observations lead us to propose that ERβ and the SRC-1 isoform act cooperatively together to affect a potent anti-apoptotic state in endometriotic tissues.

The formation of the inflammasome and the activity of caspase 1 determine the balance between pathogen resolution and disease pathology. How is the inflammasome involved in the pathogenesis of endometriosis? The NALP3 gene has an essential role in endometriosis progression because NALP3$^{-/-}$ mice have a defect in ectopic lesion growth under endometriosis. ERβ involves in upregulation of the NALP3 inflammasome in hepatocellular carcinoma cells upon estrogen stimulation even though the interaction of ERβ with NALP3 is not clearly demonstrated (Wei et al., 2015). Here, we revealed that ERβ interacts with inflammasome components and enhances inflammasome activity through the activation of caspase 1 activity. Activation of the inflammasome results in highly elevated IL-1β levels in endometriotic tissues compared with normal endometrium, and enhanced IL-1β signaling can influence the adhesion activity of endometriotic tissues and proliferative activities of human endometrial cells.

The Gain of ERβ Function May Lead to Female Infertility

One of the primary symptoms associated with endometriosis is dysfunction of the normal endometrium, leading to endometriosis-associated infertility (Holoch and Lessey, 2010). In addi-

tion to ectopic lesions, we found that eutopic endometrium demonstrated elevated levels of ERβ compared with normal endometrium. We believe that ERβ overexpression could increase endometriosis-associated infertility by preventing the decidualization response in the stromal compartment of eutopic endometrium. Thus, targeting ERβ could have dual potential benefits in patients with endometriosis: regression of ectopic lesion growth and enhancement of fecundity of women with endometriosis.

A Combination Therapy using Antagonists of ERβ and the SRC-1 Isoform Represents a Proof-of-Principle for the Next Generation of Endometriosis Therapy

Inhibitors of estrogen signaling and estrogen synthesis as well as inflammatory inhibitors (COX-2 inhibitors) have been employed, given the dependence on estrogen and the inflammatory response of ectopic lesions. However, these treatments can be associated with undesirable side effects. In addition to substantiating an infertile state in young women, long-term estrogen deficiency therapies can have harmful side effects on other estrogen target tissues, such as the brain and bone (Shah et al., 1987; Vanderschueren et al., 1997). Therefore, a greater choice of alternate therapies that more specifically target endometriotic causal modes is needed.

Our observations proposed that the targeting ERβ activity should increase the specificity and efficiency of endometriosis treatment and could be an alternative combinational approach for endometriosis treatment in lieu of current estrogen-deficiency therapy. As a proof-of-principle, the application of an ERβ-selective antagonist, such as PHTPP, significantly suppressed ectopic lesion growth by inhibiting ERβ activity in ectopic lesions of mice with endometriosis without side effects on fertility. The minimal inhibitory effects of PHTPP against uterine ERα could also be another advantage to minimize side effects. We note that a previous study stated that ERB-041, an ERβ-specific agonist, caused regression of ectopic lesion growth in an endometriosis animal model system (Harris et al., 2005). The reason for this discrepancy could potentially be related to differential ERβ expression in ectopic lesions. The expression of ERβ was not detected in human ectopic lesions that developed in athymic nude mice in Harris' study. PHTPP treatment demonstrated certain differential effects in endometriotic tissues compared with ERβ knockout tissues. For example, proliferation of eutopic endometrium and apoptosis in the stromal compartment of ectopic and eutopic endometrium were differentially regulated between them. This differential regulation might be attributable to the differences between pharmacological inhibition and genetic knockout.

Collectively, we propose that the SRC-1 isoform/ERβ complex could be a next-generation endometriosis therapeutic target with reduced side effects compared to current endometriosis treatment because (1) ERβ and the SRC-1 isoform show endometriotic tissue-specific expression but have little expression in normal endometrium; (2) both play an essential role in the early stages of endometriosis pathogenesis; and (3) targeting both of these drivers allows the marked suppression of ectopic lesion growth in animals compared with either individual agent alone.

EXPERIMENTAL PROCEDURES

Mouse Information

Five-week-old normal (C57BL/6J), ERβ$^{-/-}$ (B6;129P2-Esr2^{tm1Unc}/J), NALP3$^{-/-}$ (B6.129S6-Nlrp3^{tm1Bhk} /J), and SCID (NOD.CB17-Prkdcscid/J) mice were purchased from Jackson Laboratory. ROSA$^{LSL:ERβ/+}$ and ERBAI mice were generated. The ROSA$^{LSL:ERβ/+}$:PR$^{Cre/+}$ mice were generated by crossing ROSA$^{LSL:ERβ/+}$ with PR$^{Cre/+}$ mice (Soyal et al., 2005). All animal care was controlled by the ethical regulations approved by the Institutional Animal Care and Use Committee at Baylor College of Medicine.

Immortalized Human Endometrial Cells

iHESCs and EMosis-CC/TERT1 (immortalized human endometriotic epithelial cells) were employed and confirmed by short tandem repeat profiling; these cells were not contaminated with mycoplasma.

Surgically Induced Endometriosis

Endometriosis in mice was surgically induced under aseptic conditions under anesthesia. Details on surgically induced endometriosis are found in the Supplemental Experimental Procedures.

Generation of ERBAI and ERβ:OE Mice

Details on these mice are found in the Supplemental Experimental Procedures.

In Vivo Analysis of Human Ectopic Lesion Growth in SCID Mice

The bioluminescence images of human ectopic lesions developed with iHESCs/Luc plus iHEECs/Luc (or iHEECs/ERβ/Luc) in SCID mice were determined. Details on this are found in the Supplemental Experimental Procedures.
   For basic procedures, see the Supplemental Experimental Procedures.

Statistical Analyses

The data are expressed as the mean ± SD. Significance was assessed using an independent two-tailed Student's t test; A p value of less than 0.1 was considered statistically significant. NS, non-specific. *p < 0.1, **p < 0.01, ***p < 0.005 by Student's t test.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.034
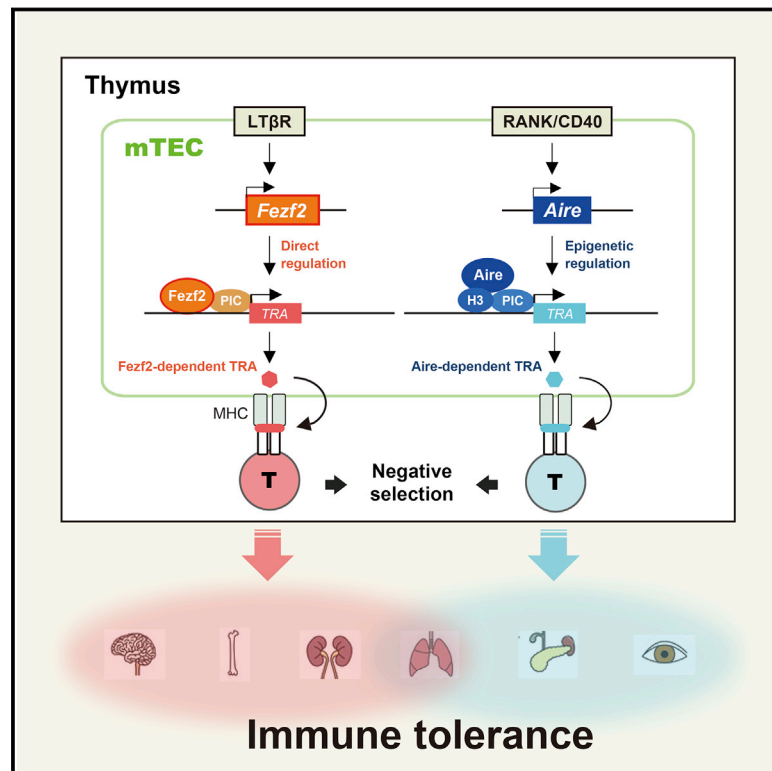
AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

REFERENCES

Acién, P., Velasco, I., Gutiérrez, M., and Martínez-Beltrán, M. (2007). Aromatase expression in endometriotic tissues and its relationship to clinical and analytical findings. Fertil. Steril. 88, 32–38.

Ahn, S.H., Monsanto, S.P., Miller, C., Singh, S.S., Thomas, R., and Tayade, C. (2015). Pathophysiology and Immune Dysfunction in Endometriosis. Biomed Res Int 2015, 795976.

Bono, Y., Kyo, S., Takakura, M., Maida, Y., Mizumoto, Y., Nakamura, M., Nomura, K., Kiyono, T., and Inoue, M. (2012). Creation of immortalised epithelial cells from ovarian endometrioma. Br. J. Cancer 106, 1205–1213.

Bratton, S.B., and Salvesen, G.S. (2010). Regulation of the Apaf-1-caspase-9 apoptosome. J. Cell Sci. 123, 3209–3214.

Bulun, S.E. (2009). Endometriosis. N. Engl. J. Med. 360, 268–279.

Bulun, S.E., Monsavais, D., Pavone, M.E., Dyson, M., Xue, Q., Attar, E., Tokunaga, H., and Su, E.J. (2012). Role of estrogen receptor-β in endometriosis. Semin. Reprod. Med. 30, 39–45.

Burney, R.O. (2013). The genetics and biochemistry of endometriosis. Curr. Opin. Obstet. Gynecol. 25, 280–286.

Burns, K.A., Rodriguez, K.F., Hewitt, S.C., Janardhan, K.S., Young, S.L., and Korach, K.S. (2012). Role of estrogen receptor signaling required for endometriosis-like lesion establishment in a mouse model. Endocrinology 153, 3960–3971.

Cao, W.G., Morin, M., Metz, C., Maheux, R., and Akoum, A. (2005). Stimulation of macrophage migration inhibitory factor expression in endometrial stromal cells by interleukin 1, beta involving the nuclear transcription factor NFkappaB. Biol. Reprod. 73, 565–570.

Cheng, C.W., Licence, D., Cook, E., Luo, F., Arends, M.J., Smith, S.K., Print, C.G., and Charnock-Jones, D.S. (2011). Activation of mutated K-ras in donor endometrial epithelium and stroma promotes lesion growth in an intact immunocompetent murine model of endometriosis. J. Pathol. 224, 261–269.

Compton, D.R., Sheng, S., Carlson, K.E., Rebacz, N.A., Lee, I.Y., Katzenellenbogen, B.S., and Katzenellenbogen, J.A. (2004). Pyrazolo[1,5-a]pyrimidines: estrogen receptor ligands possessing estrogen receptor beta antagonist activity. J. Med. Chem. 47, 5872–5893.

Delvoux, B., Groothuis, P., D'Hooghe, T., Kyama, C., Dunselman, G., and Romano, A. (2009). Increased production of 17beta-estradiol in endometriosis lesions is the result of impaired metabolism. J. Clin. Endocrinol. Metab. 94, 876–883.

Eskenazi, B., and Warner, M.L. (1997). Epidemiology of endometriosis. Obstet. Gynecol. Clin. North Am. 24, 235–258.

Giudice, L.C. (2010). Clinical practice. Endometriosis. N. Engl. J. Med. 362, 2389–2398.

Han, S.J., Hawkins, S.M., Begum, K., Jung, S.Y., Kovanci, E., Qin, J., Lydon, J.P., DeMayo, F.J., and O'Malley, B.W. (2012). A new isoform of steroid receptor coactivator-1 is crucial for pathogenic progression of endometriosis. Nat. Med. 18, 1102–1111.

Han, S.J., and O'Malley, B.W. (2014). The dynamics of nuclear receptors and nuclear receptor coregulators in the pathogenesis of endometriosis. Hum. Reprod. Update 20, 467–484.

Han, S.J., O'Malley, B.W., and DeMayo, F.J. (2009). An estrogen receptor alpha activity indicator model in mice. Genesis 47, 815–824.

Harris, H.A., Bruner-Tran, K.L., Zhang, X., Osteen, K.G., and Lyttle, C.R. (2005). A selective estrogen receptor-beta agonist causes lesion regression in an experimentally induced model of endometriosis. Hum. Reprod. 20, 936–941.

Hatai, T., Matsuzawa, A., Inoshita, S., Mochida, Y., Kuroda, T., Sakamaki, K., Kuida, K., Yonehara, S., Ichijo, H., and Takeda, K. (2000). Execution of apoptosis signal-regulating kinase 1 (ASK1)-induced apoptosis by the mitochondria-dependent caspase activation. J. Biol. Chem. 275, 26576–26581.

Holoch, K.J., and Lessey, B.A. (2010). Endometriosis and infertility. Clin. Obstet. Gynecol. 53, 429–438.

Hughesdon, P.E. (1958). Endometriosis and retrograde menstruation; case report and review. J. Obstet. Gynaecol. Br. Emp. *65*, 944–953.

Kim, J.J., Kurita, T., and Bulun, S.E. (2013). Progesterone action in endometrial cancer, endometriosis, uterine fibroids, and breast cancer. Endocr. Rev. *34*, 130–162.

Krege, J.H., Hodgin, J.B., Couse, J.F., Enmark, E., Warner, M., Mahler, J.F., Sar, M., Korach, K.S., Gustafsson, J.A., and Smithies, O. (1998). Generation and reproductive phenotypes of mice lacking estrogen receptor beta. Proc. Natl. Acad. Sci. USA *95*, 15677–15682.

Krikun, G., Mor, G., Alvero, A., Guller, S., Schatz, F., Sapi, E., Rahman, M., Caze, R., Qumsiyeh, M., and Lockwood, C.J. (2004). A novel immortalized human endometrial stromal cell line with normal progestational response. Endocrinology *145*, 2291–2296.

Miao, E.A., Rajan, J.V., and Aderem, A. (2011). Caspase-1-induced pyroptotic cell death. Immunol. Rev. *243*, 206–214.

Micheau, O., and Tschopp, J. (2003). Induction of TNF receptor I-mediated apoptosis via two sequential signaling complexes. Cell *114*, 181–190.

Monsivais, D., Dyson, M.T., Yin, P., Coon, J.S., Navarro, A., Feng, G., Malpani, S.S., Ono, M., Ercan, C.M., Wei, J.J., et al. (2014). ERβ- and prostaglandin E2-regulated pathways integrate cell proliferation via Ras-like and estrogen-regulated growth inhibitor in endometriosis. Mol. Endocrinol. *28*, 1304–1315.

Moradi, M., Parker, M., Sneddon, A., Lopez, V., and Ellwood, D. (2014). Impact of endometriosis on women's lives: a qualitative study. BMC Womens Health *14*, 123.

Nasu, K., Yuge, A., Tsuno, A., Nishida, M., and Narahara, H. (2009). Involvement of resistance to apoptosis in the pathogenesis of endometriosis. Histol. Histopathol. *24*, 1181–1192.

Ozkan, S., Murk, W., and Arici, A. (2008). Endometriosis and infertility: epidemiology and evidence-based treatments. Ann. N Y Acad. Sci. *1127*, 92–100.

Pellegrini, C., Gori, I., Achtari, C., Hornung, D., Chardonnens, E., Wunder, D., Fiche, M., and Canny, G.O. (2012). The expression of estrogen receptors as well as GREB1, c-MYC, and cyclin D1, estrogen-regulated genes implicated in proliferation, is increased in peritoneal endometriosis. Fertil. Steril. *98*, 1200–1208.

Saji, S., Jensen, E.V., Nilsson, S., Rylander, T., Warner, M., and Gustafsson, J.A. (2000). Estrogen receptors alpha and beta in the rodent mammary gland. Proc. Natl. Acad. Sci. USA *97*, 337–342.

Salmassi, A., Acar-Perk, B., Schmutzler, A.G., Koch, K., Püngel, F., Jonat, W., and Mettler, L. (2011). Apoptosis resistance in endometriosis. Bioimpacts *1*, 129–134.

Shah, A., Roberts, T., McQueen, I.N., Graham, J.G., and Walker, K. (1987). Danazol and benign intracranial hypertension. Br. Med. J. (Clin. Res. Ed.) *294*, 1323.

Sillem, M., Prifti, S., Monga, B., Arslic, T., and Runnebaum, B. (1999). Integrin-mediated adhesion of uterine endometrial cells from endometriosis patients to extracellular matrix proteins is enhanced by tumor necrosis factor alpha (TNF

alpha) and interleukin-1 (IL-1). Eur. J. Obstet. Gynecol. Reprod. Biol. *87*, 123–127.

Simoens, S., Hummelshoj, L., and D'Hooghe, T. (2007). Endometriosis: cost estimates and methodological perspective. Hum. Reprod. Update *13*, 395–404.

Soyal, S.M., Mukherjee, A., Lee, K.Y., Li, J., Li, H., DeMayo, F.J., and Lydon, J.P. (2005). Cre-mediated recombination in cell lineages that express the progesterone receptor. Genesis *41*, 58–66.

Tobiume, K., Matsuzawa, A., Takahashi, T., Nishitoh, H., Morita, K., Takeda, K., Minowa, O., Miyazono, K., Noda, T., and Ichijo, H. (2001). ASK1 is required for sustained activations of JNK/p38 MAP kinases and apoptosis. EMBO Rep. *2*, 222–228.

Tobiume, K., Saitoh, M., and Ichijo, H. (2002). Activation of apoptosis signal-regulating kinase 1 by the stress-induced activating phosphorylation of pre-formed oligomer. J. Cell. Physiol. *191*, 95–104.

Vanderschueren, D., van Herck, E., Nijs, J., Ederveen, A.G., De Coster, R., and Bouillon, R. (1997). Aromatase inhibition impairs skeletal modeling and decreases bone mineral density in growing male rats. Endocrinology *138*, 2301–2307.

Walker, V.R., and Korach, K.S. (2004). Estrogen receptor knockout mice as a model for endocrine research. ILAR J. *45*, 455–461.

Wang, Y., Lonard, D.M., Yu, Y., Chow, D.C., Palzkill, T.G., and O'Malley, B.W. (2011). Small molecule inhibition of the steroid receptor coactivators, SRC-3 and SRC-1. Mol. Endocrinol. *25*, 2041–2053.

Wei, Q., Guo, P., Mu, K., Zhang, Y., Zhao, W., Huai, W., Qiu, Y., Li, T., Ma, X., Liu, Y., et al. (2015). Estrogen suppresses hepatocellular carcinoma cells through ERβ-mediated upregulation of the NLRP3 inflammasome. Lab. Invest. *95*, 804–816.

Weigel, N.L., and Zhang, Y. (1998). Ligand-independent activation of steroid hormone receptors. J. Mol. Med. (Berl.) *76*, 469–479.

Willingham, S.B., Allen, I.C., Bergstralh, D.T., Brickey, W.J., Huang, M.T., Taxman, D.J., Duncan, J.A., and Ting, J.P. (2009). NLRP3 (NALP3, Cryopyrin) facilitates in vivo caspase-1 activation, necrosis, and HMGB1 release via inflammasome-dependent and -independent pathways. J. Immunol. *183*, 2008–2015.

Xue, Q., Lin, Z., Cheng, Y.H., Huang, C.C., Marsh, E., Yin, P., Milad, M.P., Confino, E., Reierstad, S., Innes, J., and Bulun, S.E. (2007). Promoter methylation regulates estrogen receptor 2 in human endometrium and endometriosis. Biol. Reprod. *77*, 681–687.

Zhang, H., Zhao, X., Liu, S., Li, J., Wen, Z., and Li, M. (2010). 17betaE2 promotes cell proliferation in endometriosis by decreasing PTEN via NFkappaB-dependent pathway. Mol. Cell. Endocrinol. *317*, 31–43.

Zhao, Y., Gong, P., Chen, Y., Nwachukwu, J.C., Srinivasan, S., Ko, C., Bagchi, M.K., Taylor, R.N., Korach, K.S., Nettles, K.W., et al. (2015). Dual suppression of estrogenic and inflammatory activities for targeting of endometriosis. Sci. Transl. Med. *7*, ra9.

Cell

Fezf2 Orchestrates a Thymic Program of Self-Antigen Expression for Immune Tolerance

Graphical Abstract

Highlights

- Transcription factor *Fezf2* is identified as a key regulator of autoimmunity in mTECs

- Loss of Fezf2 in mTECs leads to autoimmune disorders in mice

- Fezf2 directly regulates TRA gene expression in an Aire-independent manner

- *Fezf2* expression is regulated by the LTβR pathway but not by the RANK/CD40-Aire axis

Authors

Hiroyuki Takaba, Yasuyuki Morishita, Yoshihiko Tomofuji, ..., Noriko Komatsu, Tatsuhiko Kodama, Hiroshi Takayanagi

Correspondence

takayana@m.u-tokyo.ac.jp

In Brief

To promote immunological tolerance of one's own proteins, the protein Fezf2 directly regulates transcription of tissue-restricted antigen genes in in the thymus, where it functions independently and via a distinct pathway from Aire, the transcriptional regulator widely thought to be primarily responsible for self-tolerance.

Accession Numbers

GSE69105

CrossMark

CellPress

Fezf2 Orchestrates a Thymic Program of Self-Antigen Expression for Immune Tolerance

Hiroyuki Takaba,[1,2] Yasuyuki Morishita,[3] Yoshihiko Tomofuji,[1] Lynett Danks,[1,2] Takeshi Nitta,[1] Noriko Komatsu,[1] Tatsuhiko Kodama,[4] and Hiroshi Takayanagi[1,2,*]

[1]Department of Immunology, Graduate School of Medicine and Faculty of Medicine, The University of Tokyo, Tokyo 113-0033, Japan
[2]Takayanagi Osteonetwork Project, Exploratory Research for Advanced Technology Program, Japan Science and Technology Agency (JST), Tokyo 113-0033, Japan
[3]Department of Molecular Pathology, Graduate School of Medicine and Faculty of Medicine, The University of Tokyo, Tokyo 113-0033, Japan
[4]Research Center for Advanced Science and Technology, Department of Systems Biology and Medicine, The University of Tokyo, Tokyo 153-8904, Japan
*Correspondence: takayana@m.u-tokyo.ac.jp
http://dx.doi.org/10.1016/j.cell.2015.10.013

SUMMARY

Self-tolerance to immune reactions is established via promiscuous expression of tissue-restricted antigens (TRAs) in medullary thymic epithelial cells (mTECs), leading to the elimination of T cells that respond to self-antigens. The transcriptional regulator Aire has been thought to be sufficient for the induction of TRAs, despite some indications that other factors may promote TRA expression in the thymus. Here, we show that the transcription factor Fezf2 directly regulates various TRA genes in mTECs independently of Aire. Mice lacking Fezf2 in mTECs displayed severe autoimmune symptoms, including the production of autoantibodies and inflammatory cell infiltration targeted to peripheral organs. These responses differed from those detected in Aire-deficient mice. Furthermore, Fezf2 expression and Aire expression are regulated by distinct signaling pathways and promote the expression of different classes of proteins. Thus, two independent factors, Fezf2 and Aire, permit the expression of TRAs in the thymus to ensure immune tolerance.

INTRODUCTION

In all vertebrates, T cells are generated in the thymus, and their antigen receptors are generated through random somatic DNA rearrangement, which provides a molecular basis for recognizing and eliminating an essentially unlimited number of pathogens (Anderson and Takahama, 2012; Hogquist et al., 2005; Klein et al., 2014). However, this process inevitably gives rise to the production of self-reactive T cells. To prevent autoimmune disease, immune reactions against self-antigens must be suppressed through central and peripheral tolerance (Sakaguchi et al., 2006). The repertoire of T cell receptors (TCRs) that potently react self-components is removed in the thymus, a process that comprises the fundamental mechanism of central tolerance (Kappler et al., 1987).

T cell differentiation and repertoire selection are dependent on the thymic microenvironment, mainly supported by thymic epithelial cells (TECs) and dendritic cells (Gallegos and Bevan, 2004; Hinterberger et al., 2010). After positive selection by cortical thymic epithelial cells (cTECs) based on the proper affinity to major histocompatibility complex (MHC) molecules, T cells migrate to the medulla and undergo negative selection (Klein et al., 2014). mTECs promiscuously express peripheral tissue-restricted antigens (TRAs) (Derbinski et al., 2001) and T cells that potently react to these antigens are eliminated by apoptosis. The transcriptional regulator Aire has been shown to be essential for the promiscuous TRA expression and negative selection (Liston et al., 2003; Mathis and Benoist, 2009), since autoimmune symptoms manifest in both *Aire*-deficient (*Aire*$^{-/-}$) mice (Anderson et al., 2002) and patients with autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED) caused by mutations of the *AIRE* gene (Akirav et al., 2011). *Aire* expression is regulated by members of the tumor necrosis factor receptor superfamily such as Tnfrsf11a (also known as receptor activator of nuclear factor-κB; RANK) and CD40. In more than 10 years since the discovery of *AIRE* (Aaltonen et al., 1997; Nagamine et al., 1997), TRA expression in the thymus has been thought to be exclusively dependent on Aire. However, there are TRA genes that are induced in the absence of Aire (Derbinski et al., 2005), raising the possibility that other transcriptional regulators are functional in the induction of self-antigens in mTECs.

Here, we investigated the transcriptional regulators of TRAs in mTECs and report the identification of the key transcription factor *Fezf2* (also known as *Zfp312* and *Fezl*). Fezf2 plays an indispensable role in the expression of TRAs and the establishment of immune tolerance independently of Aire. The parallel RANK/CD40 and lymphotoxin beta receptor (LTβR) signaling pathways were shown to regulate the expression of *Aire* and *Fezf2*, respectively. Thus, these findings constitute an important advance in the current understanding of immune tolerance and the mechanisms underlying autoimmune diseases.

RESULTS

Fezf2 Is Highly Expressed in Mouse and Human mTECs

As mTECs express TRAs in the absence of Aire (Derbinski et al., 2005), we hypothesized that there are additional transcriptional
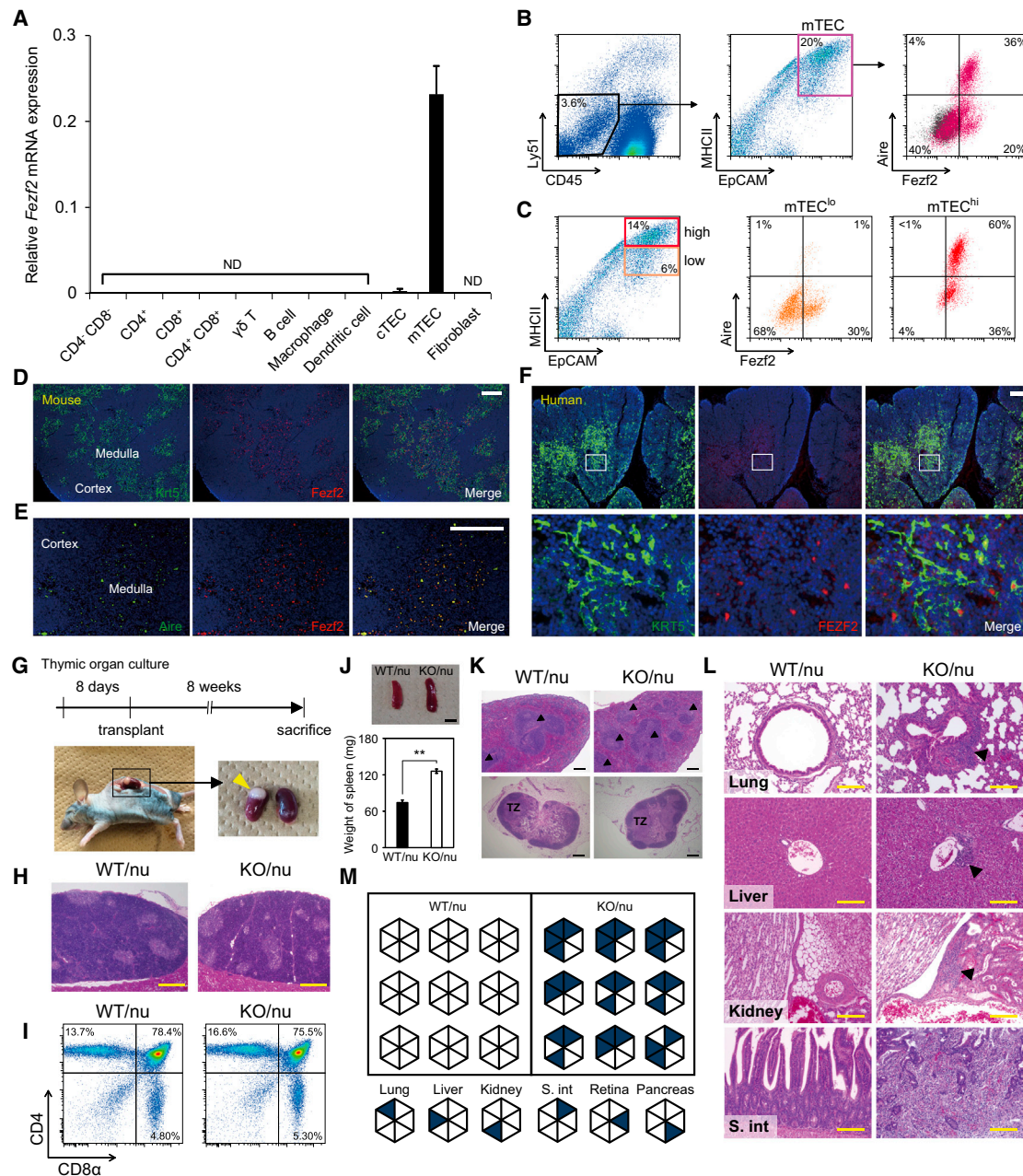
Figure 1. Loss of *Fezf2* in the Mouse mTEC Leads to Autoimmunity

(A) qRT-PCR analysis of *Fezf2* mRNA expression (relative to *Gapdh*) in various cell types in the mouse thymus (3-week-old, n = 6). Error bars indicate the mean ± SD. ND, not detected.

(B and C) Flow cytometric analysis of Fezf2 and Aire protein expression in mTECs (B) or in the mTEC^{lo} (MHC II^{low}) and mTEC^{hi} (MHC II^{high}) populations (C). mTECs were identified as EpCAM$^+$ and MHC class II$^+$ (MHC II$^+$) among the CD45$^-$ and Ly51$^-$ cells. Gray plots, the negative control.

(D and E) Immunohistochemical analysis of Fezf2 expression in keratin 5$^+$ (Krt5$^+$) mTECs (D) and Fezf2 and Aire expression in the medulla (E) of the mouse thymus.

(F) FEZF2 expression in KRT5$^+$ mTECs in human fetal thymus (29-week-old, male). The bottom panels are enlarged views of the upper panels. Scale bars in (D)–(F), 100 μm.

(G) Experimental design used for thymus transplantation (top) and macroscopic images of the thymus-grafted nude mouse (bottom left) and kidneys (bottom right). The arrowhead indicates the grafted thymus under the renal capsule.

(H) Morphological analysis of grafted thymus (H&E staining).

(I) CD4/CD8 differentiation of thymocytes in WT/nu and KO/nu mice (n = 3).

(J) Gross anatomy (top) and weight (bottom) of spleens from the WT/nu and KO/nu mice (n = 9). Scale bar, 5 mm. Data are shown as the mean ± SD. **p < 0.01.

(K) Structure of the spleen and lymph nodes of WT/nu and KO/nu mice (H&E staining). Arrowheads, follicles in the spleen; TZ, T cell zones in the inguinal lymph nodes.

(legend continued on next page)

regulators that are responsible for the induction of TRAs in the mTEC. To identify the genes selectively expressed in mTECs, we analyzed the GeneChip data on the mRNA isolated from mTECs and cTECs. Among a group of genes that were expressed 20-fold higher in mTECs than cTECs, only *Fezf2* was categorized as a transcriptional regulator (Figure S1A); thus, we focused on it as a potential mTEC-specific transcriptional regulator. Intriguingly, the expression level of *Fezf2* was as high as that of *Aire* in mTECs (Figure S1B).

Fezf2 was originally discovered in the developing forebrain of *Xenopus* and zebrafish (Eckler and Chen, 2014). Fezf2 contains the evolutionarily conserved six zinc finger domains (Figure S1C) and has been shown to be a neuronal transcription factor required for the differentiation and specification of corticospinal neurons and other subcerebral projection neurons in the brain (Shim et al., 2012), but no immunological function has been reported to date.

To examine *Fezf2* gene expression in the mouse thymus, we performed qRT-PCR analysis and showed that *Fezf2* mRNA was specifically expressed in mTECs among the various thymic cell types (Figure 1A). The expression of *Fezf2* was developmentally induced during the perinatal period (Figure S1D). The development of mTECs is characterized by the upregulation of CD80 and MHC class II (Gray et al., 2006), the expression of which determines whether mTECs belong to either the mTEC^{hi} or mTEC^{lo} population. mTEC^{hi} cells functionally produce thousands of TRAs for the negative selection of T cells (Derbinski et al., 2005; Sansom et al., 2014). Flow cytometric analysis revealed that Fezf2 is expressed in 30% of the mTEC^{lo} population and almost all of the mTEC^{hi} population (Figures 1B and 1C). Immunostaining clearly indicated that Fezf2 is expressed by mTECs positive for keratin 5, but not cTECs (Figure 1D), while almost all of the the Aire-expressing cells were positively stained for Fezf2 (Figure 1E). When we transiently overexpressed Fezf2 and Aire in an mTEC line, the Fezf2 protein specifically located in the nucleus but did not colocalize with Aire (Figure S1E). In the human thymus, immunohistochemical analysis showed that FEZF2 was selectively expressed in keratin 5-positive mTECs (Figure 1F), and AIRE was expressed in the FEZF2-positive cells (Figure S1F). These data clearly imply that Fezf2 plays a functional role in mouse and human mTECs.

Fezf2 Deficiency Leads to the Infiltration of Inflammatory Cells

Fezf2-deficient (*Fezf2$^{-/-}$*) mice have a brain development defect and fail to survive after weaning due to an inability to consume solid food (Hirata et al., 2004). The mice die at 4 weeks of age, prior to the obvious manifestation of immunological symptoms. Thus, we tested the immunological function of thymic stromal cells by transplantation of hematopoietic-cell-depleted thymuses into the renal capsule of nude mice (Anderson et al., 2002). Eight weeks later, the successful engraftment

was observed in the nude mice grafted with the *Fezf2$^{-/-}$* thymus (KO/nu) as well as in the nude mice grafted with control littermate thymus (WT/nu) (Figure 1G). There was no difference in *Aire* expression (Figure S1G) or thymic size between the KO/nu and WT/nu mice (Figure 1H). Flow cytometric analysis revealed a normal differentiation of CD4 and CD8 thymocytes in the engrafted thymuses (Figure 1I), while the number of Foxp3$^+$ regulatory T (Treg) cells was slightly decreased in KO/nu mice (Figure S1H). KO/nu mice exhibited splenomegaly with an increased number of follicles as well as an expansion of the T cell zone in the lymph nodes (Figures 1J and 1K). An infiltration of inflammatory cells, including lymphocytes, was observed in peripheral tissues, including the lung, liver, kidney, and small intestine (Figures 1L and 1M). However, there was no obvious inflammatory cell infiltration into the pancreas or retina (Figure S1I), which are highly affected in *Aire$^{-/-}$* mice (Anderson et al., 2002).

Thymocyte and mTEC Development in *Fezf2$^{-/-}$* Mice

To examine the function of Fezf2 in the development of mTECs and thymocytes, *Fezf2$^{-/-}$* mice were analyzed before weaning, at the 3 weeks of age. Keratin 5-positive mTECs were abnormally distributed in clusters in the thymic medulla of the *Fezf2$^{-/-}$* mice as compared to the wild-type (WT) mice (Figures 2A, S2A, and S2B). The cTEC number was unchanged (Figure S2C), but the mTEC number and the ratio of mTECs (CD80$^+$ cells) to total TECs (EpCAM$^+$CD45$^-$ cells) were decreased in *Fezf2$^{-/-}$* mice (Figures 2B, 2C, and S2D). However, there was no significant difference in the ratio of mTEC^{hi} and mTEC^{lo} (Figures 2D and S2E) or in Aire protein as well as mRNA expression in mTECs (Figures 2E and 2F) between WT and *Fezf2$^{-/-}$* mice. Collectively, these results suggest that *Fezf2* is not involved in the regulation of *Aire* expression in mTECs or the mTEC$^{hi/lo}$ ratio, although thymic organization and the mTEC number are slightly influenced by *Fezf2* deficiency. The differentiation of thymocytes into CD4$^+$ or CD8$^+$ T cells was normal (Figure 2G), while the ratio of Foxp3$^+$ to CD4$^+$ thymocytes was modestly decreased in *Fezf2$^{-/-}$* mice (Figure 2H). We found that *Fezf2$^{-/-}$* mice exhibited shifts in the CD4$^+$ and CD8$^+$ TCRVβ repertoire in the thymus (Figure 2I), suggesting a crucial role of Fezf2 in shaping the TCR repertoire of CD4$^+$ and CD8$^+$ T cells.

Fezf2 Controls TRA Expression Independently of Aire

How does Fezf2 contribute to thymic gene expression? We performed a genome-wide analysis of mRNAs expressed in the mTECs isolated from *Fezf2$^{-/-}$* mice and selected genes that were expressed more than four times higher in the WT than the *Fezf2$^{-/-}$* mTECs (Figure 3A). All of the 16 Fezf2-dependent genes belonged to TRAs, which are expressed in specific peripheral tissues according to the BioGPS gene annotation portal (http://biogps.org/#goto=welcome), and are thus regarded as Fezf2-dependent TRAs (Table S1). Interestingly, five of these Fezf2-dependent TRAs (*Krt10*, *Resp18*, *Fabp9*, *Maoa*, and

(L) Inflammatory cell infiltration in the lung, liver, kidney (arrowheads), and small intestine (S. int) of the KO/nu mice (H&E staining).

(M) Schematic of inflammatory cell infiltration into the peripheral tissues of WT/nu and KO/nu mice. Each hexagon represents a single grafted mouse (n = 9). The blue triangle indicates the detection of inflammatory cell infiltration in the organ. Scale bars in (H), (K), and (L), 200 μm.
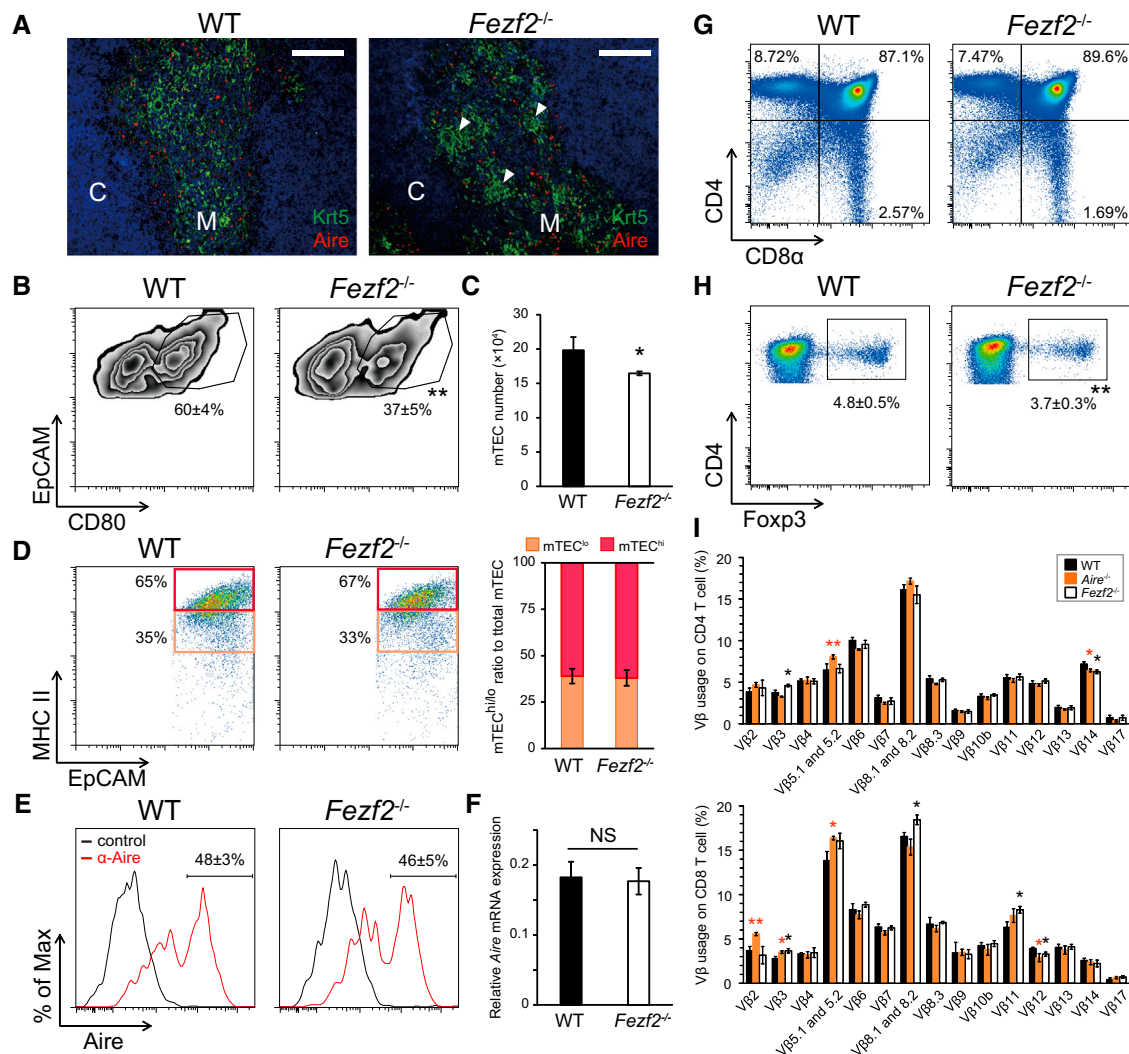
See also Figure S1.

Figure 2. Fezf2 Shapes the Repertoire of T Cell Receptors

(A) Distribution of mTECs and Aire protein expression in the WT and *Fezf2*$^{-/-}$ thymus. C and M represent the cortex and medulla, respectively, and the arrowheads indicate clustered Krt5-positive mTECs. Scale bars, 100 μm.

(B) The ratio of the total mTECs (CD80$^+$ cells) to TECs (CD45$^-$EpCAM$^+$ cells) (WT, n = 8; *Fezf2*$^{-/-}$, n = 9).

(C) The total mTEC number in WT and *Fezf2*$^{-/-}$ mice (n = 8 per genotype).

(D) The ratio of mTEC^{hi} and mTEC^{lo} (MHCII^{hi} and MHCII^{lo}) among the total mTECs (CD45$^-$Ly51$^-$MHCII$^+$EpCAM$^+$ cells) (WT, n = 3; *Fezf2*$^{-/-}$, n = 4).

(E) Aire protein expression in mTECs in WT and *Fezf2*$^{-/-}$ mice. The numbers above the lines indicate the ratio of Aire$^+$ cells to the total mTECs (WT, n = 6; *Fezf2*$^{-/-}$, n = 8).

(F) *Aire* mRNA expression in mTECs (WT, n = 10; *Fezf2*$^{-/-}$, n = 8).

(G) The normal differentiation of thymocytes into CD4$^+$ or CD8$^+$ T cells (WT, n = 10; *Fezf2*$^{-/-}$, n = 8).

(H) The ratio of Foxp3$^+$ cells to CD4$^+$ thymocytes (WT, n = 6; *Fezf2*$^{-/-}$, n = 5).

(I) The usage of TCRVβ on CD4$^+$ or CD8$^+$ T cells in the thymus (WT, n = 10; *Fezf2*$^{-/-}$, n = 8; *Aire*$^{-/-}$, n = 5). Data are shown as the mean ± SD. NS, not significant (p > 0.05); *p < 0.05; **p < 0.01.

See also Figure S2.

Timd2) had already been reported to be induced in the absence of Aire, i.e., Aire-independent TRA genes (Derbinski et al., 2005). These data suggest that Fezf2 regulates the expression of a distinct subset of TRAs in the thymus. On the other hand, there were no significant differences in the expression of MHC class I and II molecules (*H2-D*, *H2-K*, and *H2-A*, *H2-E,* respectively), costimulatory molecules (*Cd80* and *Cd86*), invariant chain

(*Cd74*), chemokines (*Xcl1* [Lei et al., 2011] and *Ccl21* [Laan et al., 2009]) or mTEC terminal differentiation marker *Ivl* (Nishikawa et al., 2010) (Figures S3A–S3E).

qRT-PCR analysis revealed that mRNA expression of the Fezf2-dependent TRAs was not reduced in the mTECs isolated from *Aire*$^{-/-}$ mice (Figure 3B). The expression of representative Aire-dependent genes (*Spt1*, *Ins2*, *Mup4*, and *S100a8*)
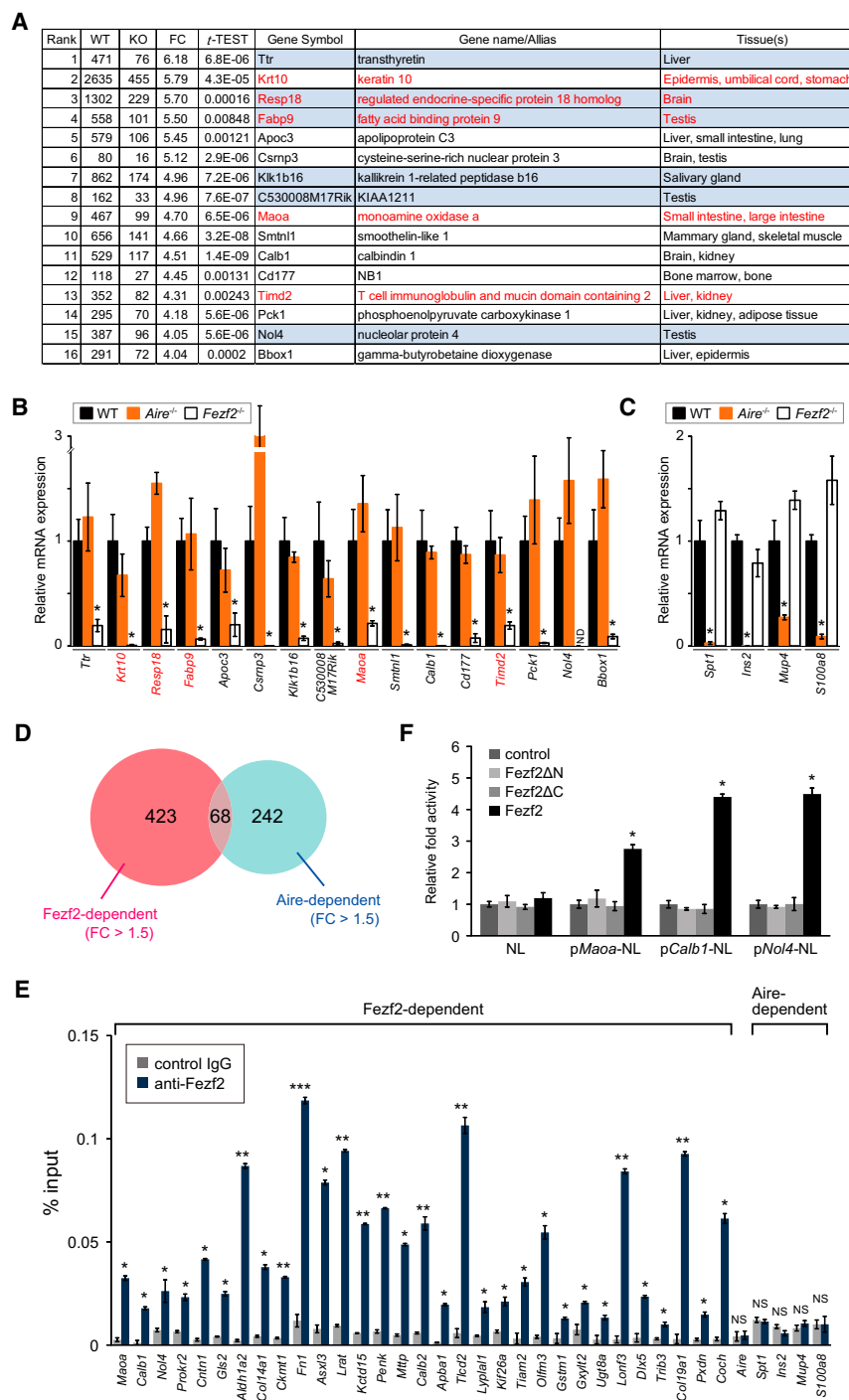
A

Rank	WT	KO	FC	t-TEST	Gene Symbol	Gene name/Allias	Tissue(s)
1	471	76	6.18	6.8E-06	Ttr	transthyretin	Liver
2	2635	455	5.79	4.3E-05	Krt10	keratin 10	Epidermis, umbilical cord, stomach
3	1302	229	5.70	0.00016	Resp18	regulated endocrine-specific protein 18 homolog	Brain
4	558	101	5.50	0.00848	Fabp9	fatty acid binding protein 9	Testis
5	579	106	5.45	0.00121	Apoc3	apolipoprotein C3	Liver, small intestine, lung
6	80	16	5.12	2.9E-06	Csrnp3	cysteine-serine-rich nuclear protein 3	Brain, testis
7	862	174	4.96	7.2E-06	Klk1b16	kallikrein 1-related peptidase b16	Salivary gland
8	162	33	4.96	7.6E-07	C530008M17Rik	KIAA1211	Testis
9	467	99	4.70	6.5E-06	Maoa	monoamine oxidase a	Small intestine, large intestine
10	656	141	4.66	3.2E-08	Smtnl1	smoothelin-like 1	Mammary gland, skeletal muscle
11	529	117	4.51	1.4E-09	Calb1	calbindin 1	Brain, kidney
12	118	27	4.45	0.00131	Cd177	NB1	Bone marrow, bone
13	352	82	4.31	0.00243	Timd2	T cell immunoglobulin and mucin domain containing 2	Liver, kidney
14	295	70	4.18	5.6E-06	Pck1	phosphoenolpyruvate carboxykinase 1	Liver, kidney, adipose tissue
15	387	96	4.05	5.6E-06	Nol4	nucleolar protein 4	Testis
16	291	72	4.04	0.0002	Bbox1	gamma-butyrobetaine dioxygenase	Liver, epidermis

Figure 3. Fezf2 Regulates a Subset of TRA Genes in an Aire-Independent Manner

(A) Fezf2-dependent genes in mTEC based on mRNA expression profiling of mTECs from WT and $Fezf2^{-/-}$ (KO) mice (n = 5 per genotype). Listed are the 16 most highly downregulated genes in $Fezf2^{-/-}$ mTECs compared with WT mTECs. The genes that have been reported as Aire-independent TRA genes are shown in red. Genes expressed in a single tissue are highlighted in blue shading.

(B) Comparison of mRNA expression of Fezf2-dependent TRA genes in WT, $Aire^{-/-}$, and $Fezf2^{-/-}$ mTECs.

(C) The mRNA expression of representative Aire-dependent TRA genes in WT, $Aire^{-/-}$, and $Fezf2^{-/-}$ mTECs. The data in (B) and (C) are shown as the mean ± SD, *p < 0.05 versus WT (WT, n = 4; $Fezf2^{-/-}$, n = 3; $Aire^{-/-}$, n = 5); ND, not detected.

(D) Chromatin immunoprecipitation (ChIP) assay. Fezf2 binds to the promoter region of Fezf2-dependent but not Aire or Aire-dependent TRA genes.

(E) A comparison of the Fezf2-dependent and Aire-dependent genes. Most of the Fezf2-dependent genes are different from the Aire-dependent genes collected from the database (GeneChip data: GSE69105 and GSE85; fold change [FC] >1.5, p < 0.05).

(F) Luciferase assay in an mTEC line 1C6. The expression of Fezf2-dependent TRA genes (Maoa, Calb1, and Nol4) is controlled by Fezf2, but not by N-terminus-deleted Fezf2 (Fezf2ΔN) or C-terminus-deleted Fezf2 (Fezf2ΔC). Luciferase vectors: empty pNluc vector (NL) or pNluc vectors containing the promoter region of Maoa, Calb1, or Nol4. Data in (D) and (F) are shown as the mean ± SD, *p < 0.05, **p < 0.01, ***p < 0.001. See also Figure S3.

ure 1C). Bioinformatics analysis revealed that Aire and/or Fezf2 control over 60% of the mTEC-specific TRAs (Figure S3G). The genes downregulated in $Fezf2^{-/-}$ mTECs included certain TRAs related to autoimmune or neoplastic diseases (Endo et al., 2009; Fatourou and Koskinas, 2009; Roulois et al., 2013) (Figure S3H). These results show that Fezf2 regulates a unique subset of TRA genes independently of Aire.

A recent global chromatin immunoprecipitation sequencing (ChIP-seq) analysis revealed that Fezf2 binds to the promoter

(Waterfield et al., 2014) was not significantly decreased in $Fezf2^{-/-}$ mice (Figure 3C). In addition, microarray analysis revealed that the downregulated genes in the $Fezf2^{-/-}$ mTECs were for the most part different from those downregulated in $Aire^{-/-}$ (Figure 3D). The expression of most of the Fezf2-dependent genes was higher in mTEC^{hi} than mTEC^{lo} (Figure S3F), consistent with the higher expression of Fezf2 in mTEC^{hi} (Fig-

of numerous protein-coding genes in cortical progenitors (Lodato et al., 2014). Our ChIP assay demonstrated that Fezf2 directly bound to the promoter region of Fezf2-depenedent TRA genes, but not to that of the Aire or Aire-dependent TRA genes in mTECs (Figure 3E). Additionally, we found Fezf2 bound to the promoter region of certain TRA genes such as Mbp, Gad1, Col2a1, and Muc1, which are known to be autoantigens or tumor
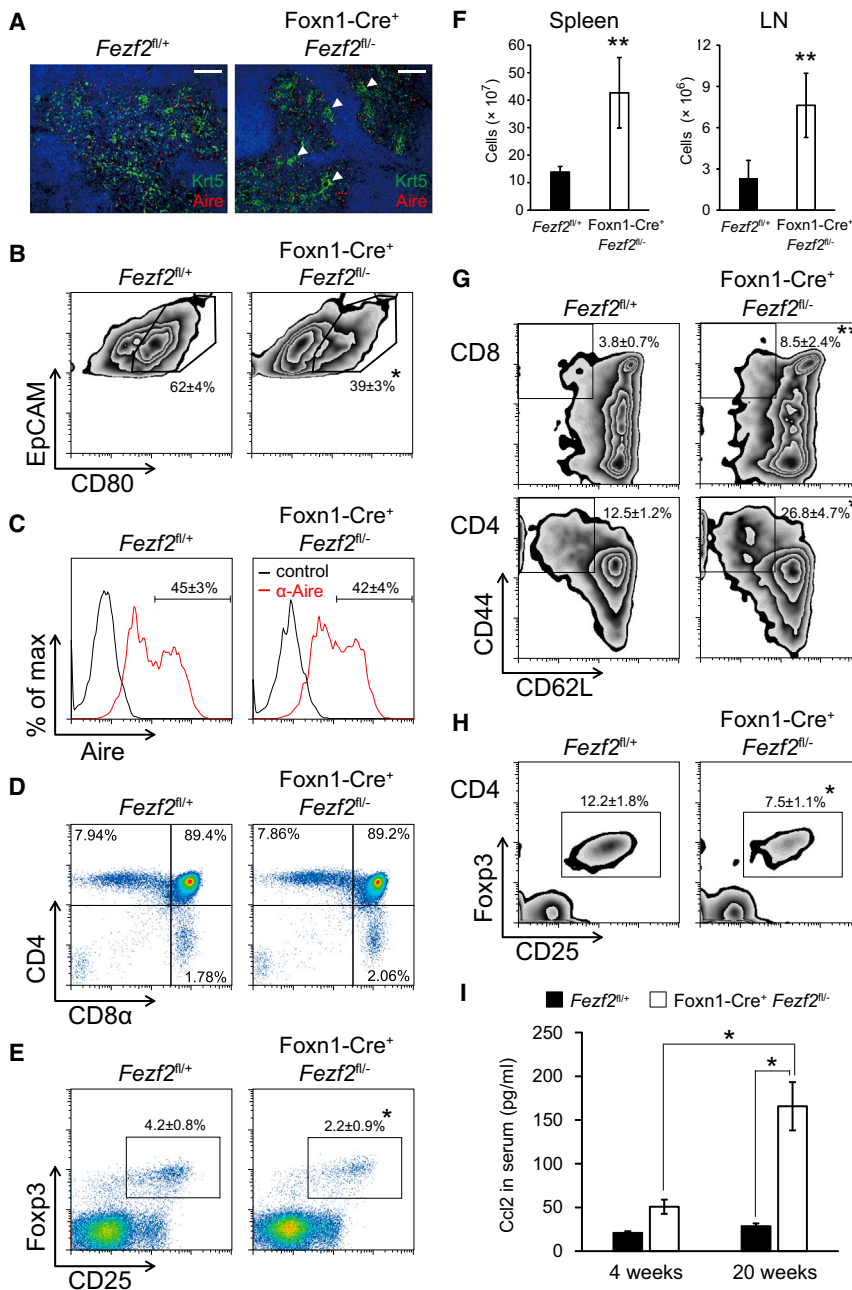
Figure 4. Analyses of Lymphoid Organs in TEC-Specific *Fezf2*-Deficient Mice

(A) Localization of mTECs in the thymic medulla in Foxn1-Cre[+] *Fezf2*[fl/−] mice. The arrowhead indicates the clustered mTECs. Scale bars, 100 μm.

(B) The ratio of mTECs (CD45[−], EpCAM[+], and CD80[+] cells) to the total TECs (n = 8 per genotype).

(C) Aire protein expression in mTECs (n = 6 per genotype).

(D) CD4/CD8 differentiation of thymocytes (n = 12 per genotype).

(E) The ratio of Foxp3[+]CD25[+] CD4 T cells in the thymus (n = 12 per genotype).

(F) The increased number of T cells in spleen and lymph nodes (LN) in Foxn1-Cre[+] *Fezf2*[fl/−] mice (n = 4 per genotype).

(G) The elevated ratio of CD62L[lo]CD44[hi] effector/memory T cells in Foxn1-Cre[+] Fezf2[fl/−] mice (n = 6 per genotype).

(H) The reduced ratio of Foxp3[+]CD25[+] CD4 T cells in the LN of Foxn1-Cre[+] *Fezf2*[fl/−] mice (n = 6 per genotype).

(I) The serum Ccl2 (MCP-1) level measured by cytometry beads assay (n = 5 per genotype). Data are shown as the mean ± SD. *p < 0.05; **p < 0.01.

See also Figure S4.

immunity. In accordance with the findings in the *Fezf2*[−/−] mice, keratin 5-positive mTECs were also abnormally distributed in clusters in the thymic medulla of Foxn1-Cre[+] *Fezf2*[fl/−] mice (Figure 4A). Both the mTEC number and the fraction of mTECs among the total TECs were decreased (Figures 4B and S4A). However, the expression of Aire in the mTECs (Figure 4C) and the ratio of mTEC[hi] and mTEC[lo] (Figure S4B) were unaffected in Foxn1-Cre[+] *Fezf2*[fl/−] mice. The differentiation of thymocytes into CD4[+] or CD8[+] T cells in the thymus was unaltered (Figure 4D). B cells differentiated normally in the bone marrow, spleen, and lymph nodes (Figures S4C–S4E). There was a modest reduction in the ratio of Foxp3[+]CD25[+] to

antigens (Figure S3I) (Ali et al., 2011; Belogurov et al., 2008; Komatsu et al., 2014; Roulois et al., 2013). Luciferase reporter assay revealed that Fezf2 regulates the expression of *Maoa*, *Calb1*, and *Nol4* genes (Figure 3F). Therefore, these data indicate that Fezf2 directly binds to the promoter region and controls the gene expression of distinct TRAs in an Aire-independent manner.

The Generation and Immunological Analysis of TEC-Specific *Fezf2*-Deficient Mice

We generated the TEC-specific *Fezf2*-deficient (Foxn1-Cre[+] *Fezf2*[fl/−]) mice and further examined the role of Fezf2 in auto-

CD4[+] cells in the Foxn1-Cre[+] *Fezf2*[fl/−] thymus (Figures 4E and S4F). The percentage of CD4[+] and CD8[+] T cells was unchanged (Figure S4G), but the T cell number was significantly elevated in the spleen and lymph nodes of the Foxn1-Cre[+] *Fezf2*[fl/−] mice (Figure 4F). The Foxn1-Cre[+] *Fezf2*[fl/−] mice showed an elevated proportion of CD44[hi]CD62L[−] effector/memory T cells (Figure 4G), while the percentages of Foxp3[+]CD25[+] Treg cells were reduced in the lymph nodes (Figure 4H). The Foxn1-Cre[+] *Fezf2*[fl/−] mice had a higher serum level of Ccl2 (also known as Mcp-1) (Figure 4I), a chemokine that induces the migration and infiltration of monocytes and T cells (Deshmane et al., 2009).

TEC-Specific *Fezf2*-Deficient Mice Display Autoimmune Symptoms, Including Autoantibody Production and Peripheral Organ Inflammation

We analyzed the infiltration of inflammatory cells, including lymphocytes, into peripheral tissues, as well as autoantibodies in the sera of Foxn1-Cre[+] *Fezf2*[fl/−] mice. Inflammatory cell infiltration was detected in the lung, liver, kidney, stomach, small intestine, salivary gland, brain, and testis of Foxn1-Cre[+] *Fezf2*[fl/−] mice (Figure 5A). Thus, the spectrum of autoimmune target tissues was found to be wider than that of *Aire*[−/−] mice. There was, however, no obvious inflammatory cell infiltration into the pancreas or retina, which is consistent with the observation in KO/nu mice (Figure S1I). Foxn1-Cre[+] *Fezf2*[fl/−] mice exhibited hypergammaglobulinemia (Figure S5). To assess whether the sera from KO/nu or Foxn1-Cre[+] *Fezf2*[fl/−] mice contained autoantibodies, we immunostained various tissues derived from *Rag1*[−/−] mice with the sera. Antibody reactivity was detected in both KO/nu and Foxn1-Cre[+] *Fezf2*[fl/−] sera against the lung, retina, skin, and joint tissues, but the staining patterns were different from that of *Aire*[−/−] serum (Figure 5B), suggesting that *Fezf2*[−/−] and *Aire*[−/−] mice have autoantibodies recognizing distinct tissue antigens. The staining pattern of the serum from Foxn1-Cre[+] *Fezf2*[fl/−] mice was much broader than that of *Aire*[−/−]. Approximately 30% of the TEC-specific *Fezf2*-deficient mice (five of 16) died by 12 months, whereas all of the *Aire*[−/−] mice survived for more than 16 months after birth (20/20), implying the autoimmune symptoms in the case of *Fezf2* deficiency were more severe than that in *Aire*[−/−] deficiency. The serum from Foxn1-Cre[+] *Fezf2*[fl/−] mice contained significantly increased autoantibodies that recognize recombinant proteins of the Fezf2-dependent TRAs, Ttr, Maoa, and Clab1, as compared to the sera from WT and *Aire*[−/−] mice (Figure 5C). In contrast, the serum from the Foxn1-Cre[+] *Fezf2*[fl/−] mice was not reactive to Aire-dependent TRAs, such as Irbp (DeVoss et al., 2006) (Figure 5C). Taken together, we conclude that Fezf2 functions in mTECs to ensure immunological tolerance against certain tissue antigens and effectively suppresses the development of autoimmune responses.

Fezf2 Is Regulated by the LTβR Pathway, but Not by the RANK/CD40 Pathways or Aire

The receptors of the TNF superfamily are required for the formation of the thymic microenvironment and central tolerance. For example, the RANK and CD40 signaling pathways are essential for mTEC development and *Aire* expression (Akiyama et al., 2008). Therefore, we examined the expression of *Fezf2* in mice deficient in *Tnfrsf11* (encoding *RANK*) and *Cd40*. Fezf2-expressing cells were normally observed upon immunohistochemical analysis and *Fezf2* mRNA expression was unaffected in Foxn1-Cre[+] *Tnfrsf11a*[fl/−] mice (Figures S6A–S6C) as well as *Cd40*[−/−] mice (Figures S6D–S6F). Normal *Fezf2* expression was also observed in mice deficient in *Tnfsf11* (encoding *RANKL*) (Figures 6A and 6B). To examine the possibility that Aire controls *Fezf2* expression, we analyzed *Aire*[−/−] mice and found no difference in *Fezf2* expression between the WT and *Aire*[−/−] mice (Figures 6C–6E). These results indicate that *Fezf2* expression is not regulated by either the RANK/CD40 pathways or Aire.

Both TNF-receptor-associated factor 6 (Traf6) and LTβR signaling critically contribute to mTEC differentiation in the establishment of the medullary microenvironment and central tolerance (Akiyama et al., 2005; Boehm et al., 2003). *Traf6*[−/−] mice displayed a significant reduction in *Fezf2* mRNA and protein expression in the thymus (Figures S6G and S6H). The ratio of mTEC to the total TECs was significantly decreased in *Ltbr*[−/−] mice (Figure 6F), and *Fezf2* mRNA and protein expression was decreased in the *Ltbr*[−/−] mTECs (Figures 6G–6I). Consistent with this, the expression of the Fezf2-dependent TRA genes *Krt10*, *Fabp9*, *Calb1*, and *Nol4* were markedly decreased (Figure 6J), while the expression of Aire-dependent TRAs as well as *Aire* itself were unaffected in the *Ltbr*[−/−] mTECs (Martins et al., 2008; Venanzi et al., 2007). Thus, *Fezf2* expression is induced by the LTβR pathway independently of the RANK/CD40 pathways or Aire.

DISCUSSION

The Discovery of a Key Transcription Factor Involved in the Establishment of Central Tolerance

In this study, it has been demonstrated that Fezf2, which is exclusively expressed in mTECs in the thymus, regulates TRA expression independently of Aire. Mice specifically lacking *Fezf2* in mTECs develop severe autoimmune phenotypes, with autoantibody production and inflammatory cell infiltration. Notably, Fezf2 and Aire play non-redundant roles in the TRA expression that is crucial for the elimination of self-reactive T cells (Figure 7). Thus, Fezf2 functions in the thymus as an mTEC-specific transcription factor required for the establishment of central tolerance, thereby playing a key role in the regulation of untoward immunological reactions.

Distinct Function of the RANK/CD40 and LTβR Pathways in the Expression of TRAs in mTECs

The TNF receptors RANK, CD40, and LTβR are required for central tolerance and TRA expression (Akiyama et al., 2008; Martins et al., 2008). The RANK and CD40 signaling pathways regulate *Aire* expression and the development of mTECs, and *Spi-B* is involved in the differentiation of mTEC[lo] to mTEC[hi] downstream of the RANK signaling (Akiyama et al., 2014). The LTβR signaling pathway is needed for the formation of normal thymic architecture and mTEC differentiation (Boehm et al., 2003; Venanzi et al., 2007; White et al., 2010), but *Ltbr* deficiency does not affect either *Aire* expression or Aire-dependent TRA gene expression in mTECs (Seach et al., 2008). The evidence obtained in this study shows that *Fezf2* is a transcriptional regulator that is induced downstream of the LTβR signaling, and that Fezf2 regulates the expression of certain genes (Figure 3) but is not involved in the differentiation of mTEC[lo] to mTEC[hi] (Figures 2D and S4B). Fezf2-dependent TRA expression, but not Aire-dependent TRA expression, was decreased in *Ltbr*[−/−] mTECs (Figure 6J). Thus, RANK/CD40 and LTβR regulate the induction of distinct TRA genes through Aire and Fezf2, respectively.

It will be of considerable interest to investigate why there are two pathways for the establishment of central tolerance in mTECs. The LTβR signaling pathway is essential for lymphoid
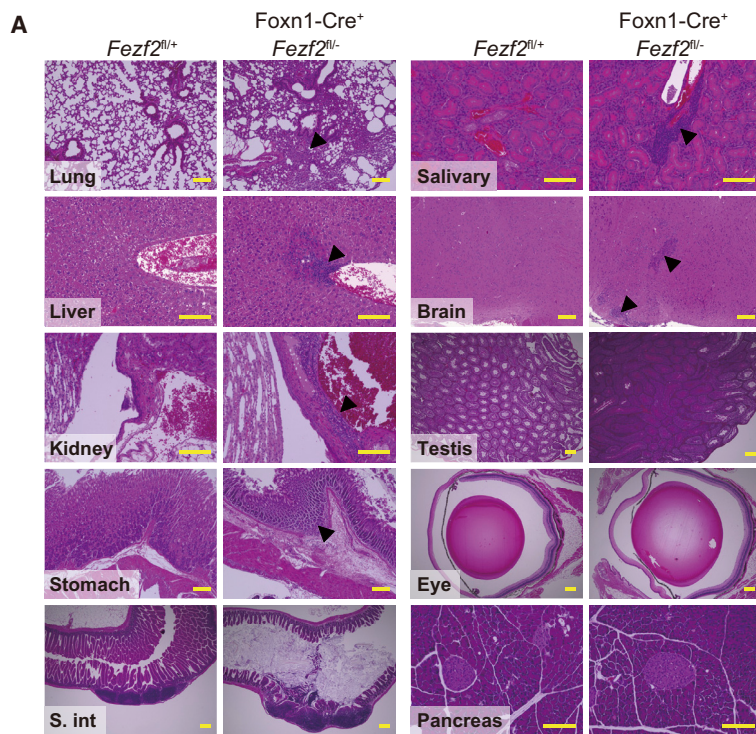
Figure 5. Tissue Infiltration and Autoantibody Production in TEC-Specific *Fezf2*-Deficient Mice

(A) Inflammatory cell infiltration in the peripheral tissues in the Foxn1-Cre[+] *Fezf2*[fl/−] mice. Infiltration was detected in lung, liver, kidney (arrowheads), and small intestine (S. int), but not in the retina or pancreas (H&E staining).

(B) Detection of autoantibody reactivity in the serum of Foxn1-Cre[+] *Fezf2*[fl/−] mice. Sections of tissues from *Rag1*[−/−] mice were stained with sera from WT/nu, KO/nu, *Aire*[−/−], or Foxn1-Cre[+] *Fezf2*[fl/−] mice followed by a secondary antibody labeled with Alexa 555. Serum was collected from *Aire*[−/−] or Foxn1-Cre[+] *Fezf2*[fl/−] mice at the age of 20 weeks. Scale bars in (A) and (B), 200 μm.

(C) Detection of autoantibodies against Fezf2-dependent (Ttr, Maoa, and Calb1) or Aire-dependent (Irbp) TRAs in the serum from Foxn1-Cre[+] *Fezf2*[fl/−] mice (ELISA, n = 6 per genotype). Data are shown as the mean ± SD. *$p < 0.05$; **$p < 0.01$.

See also Figure S5.
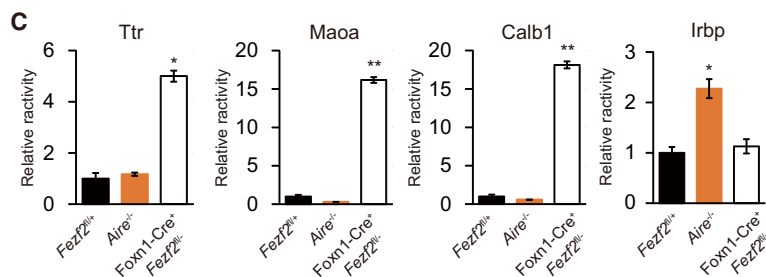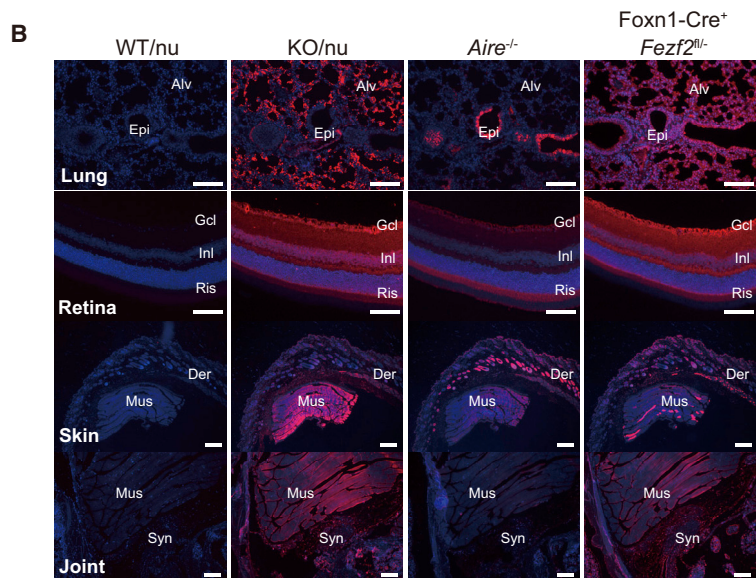
organogenesis, while the RANK and CD40 signaling pathways are required for bone homeostasis and the T-cell-mediated B cell response, respectively (Locksley et al., 2001). Although the *Fezf2* gene is conserved in all vertebrates (Eckler and Chen, 2014) that have lymphoid organs (Hofmann et al., 2010; Takaba et al., 2013), the *Aire* gene is conserved only in jawed vertebrates (Saltis et al., 2008). During the course of evolution, the LTβR-Fezf2 axis might have emerged as the primitive pathway for immune tolerance, while the RANK/CD40-Aire axis emerged later to perform certain critical functions not carried out by Fezf2.

Direct Transcriptional Regulation of TRA Genes by Fezf2

Aire does not have an obvious DNA binding domain and is not regarded as authentic transcription factor (Abramson et al., 2010; Giraud et al., 2014). Aire interacts with several nuclear factors such as hypomethyrated histone H3 (Koh et al., 2010) and the ATF7ip-MBD1 protein complex (Waterfield et al., 2014), unleashing the stalled RNA polymerases in order to epigenetically promote the ectopic transcription of TRA genes (Peterson et al., 2008). In contrast, Fezf2 has a DNA binding domain and directly regulates the expression of a large number of genes (Eckler and Chen, 2014). Although a recent global ChIP-seq analysis showed that Fezf2 binds to the promoters of a large number (over 10,000) of protein-coding genes (Lodato et al., 2014), Fezf2 does not bind the promoters of *Aire* or Aire-dependent TRA genes, such as *Ins2* (Figures 3D and S3I). While Aire expression is mainly limited to mTECs, Fezf2 is expressed and plays a role in other organs, such as the brain. Further studies are necessary to understand how Fezf2 functions as an inducer of TRA in mTECs. One of the possibilities is that Fezf2 might interact with mTEC-specific transcriptional regulator(s) that specifically modulate chromatin structure and mediate the tissue-specific function of Fezf2.

By our bioinformatics analysis, most Fezf2-dependent genes did not overlap with Aire-dependent genes (Figure 3D). Although some TRA genes might be regulated by Aire and Fezf2 co-operatively, the majority of TRAs were found to be separately controlled by either Fezf2 or Aire in mTECs (Figure S3G). Since the expression of certain TRAs was unaffected in both $Aire^{-/-}$ and $Fezf2^{-/-}$ mice, we cannot rule out the possibility that another transcriptional regulator(s) contributes to the TRA expression in mTECs.

The Spectrum of Autoimmunity in *Fezf2*-Deficient Mice

In *Fezf2* deficiency, inflammatory cell infiltration was observed in several peripheral tissues, but not in the retina or pancreas, which are often affected in *Aire* deficiency (Figure S1I) (Anderson et al., 2002). Antibodies in the sera from KO/nu mice recognized peripheral tissues of *Rag1*-deficient mice, and the staining pattern was largely complementary to that of $Aire^{-/-}$ mice (Figure 5B). In addition, the serum from Foxn1-Cre[+] $Fezf2^{fl/-}$ mice contained significantly increased autoantibodies that recognize Fezf2-dependent but not Aire-dependent TRAs (Figure 5B; Table S2). Fezf2 directly binds to the promoter region of Fezf2-dependent TRA genes in the mTECs, but not to that of but not Aire-dependent TRA genes (Figure 3E). These

data strongly support the conclusion that the *Fezf2*-deficient mice have autoimmune reactions in a selective and antigen-dependent manner, and the failure of clonal deletion causes the autoimmune phenotypes in *Fezf2* deficiency, even though they exhibited small reductions in the mTEC and thymic Treg cell number and there was disorganization of mTECs in the medulla. However, since Aire is also involved in mTEC and thymic Treg cell development (Anderson and Takahama, 2012; Yang et al., 2015), it will be important in future studies to investigate how Fezf2 contributes to mTEC development and T cell selection, including clonal deletion, anergy, and Treg cell differentiation in the thymus.

Association of Fezf2-Dependent TRAs and *FEZF2* Mutations with Autoimmune and Neoplastic Diseases

Fezf2-dependent genes included certain TRAs related to autoimmune diseases or neoplastic diseases (Figures 3A and S3G): *Ttr* is related to an autoantibody in rheumatoid arthritis (Sharma et al., 2014), *Amy2a* is involved in autoimmune pancreatitis and fulminant type 1 diabetes (Endo et al., 2009), and *Afp* and *Muc1* are reported to be tumor antigens in several different cancers (Fatourou and Koskinas, 2009; Roulois et al., 2013). Intriguingly, Fezf2-dependent TRAs are mostly categorized as intracellular or plasma membrane proteins. In contrast, many Aire-dependent TRAs are secretory proteins (Table S1), which is consistent with the fact that patients with APECED are afflicted with endocrine disorders.

It was demonstrated that FEZF2 is highly and specifically expressed in mTECs among the various types of thymic cells in humans. Although there is no report in the literature on autoimmune diseases caused by FEZF2 mutation, certain Fezf2-dependent TRAs are related to autoimmune diseases and cancers (Endo et al., 2009; Fatourou and Koskinas, 2009; Roulois et al., 2013; Sharma et al., 2014) and *FEZF2* mutations have been associated with autism (Kwan, 2013; Sanders et al., 2012) and tumors (Shu et al., 2013). This study has revealed the transcriptional program in the thymus governed by Fezf2 and Aire, both of which contribute in an essential manner to the suppression of autoimmune diseases. Thus, this study represents an important advance in our understanding of the molecular mechanisms underlying the establishment of central tolerance and the adaptive immune system.

EXPERIMENTAL PROCEDURES

Mice

$Fezf2^{+/-}$ (CDB0498K) (Hirata et al., 2004) and $Ltbr^{+/-}$ (CDB0531K) (Mouri et al., 2011) mice were obtained from the animal facility of the Center for Developmental Biology, Riken, Japan (http://www2.clst.riken.jp/arg/mutantmicelist.html). $Aire^{+/-}$ (stock number, #004743), Foxn1-Cre (#018448), and nude mice (#000819) were purchased from the Jackson Laboratory. $Fezf2^{fl/fl}$ mice (Han et al., 2011), $Tnfsf11a^{-/-}$ mice, and $Rag1^{-/-}$ mice were described previously. Mice were maintained under specific pathogen-free conditions and handled in accordance with the guidelines for animal experiments of the University of Tokyo.

Flow Cytometric Analysis and Cell Isolation

The following fluorescence-conjugated antibodies were used in the flow cytometric analysis and cell sorting. Flow cytometric analysis was performed with a FACSCanto II (BD Biosciences) and FlowJo software (Tree Star). Thymic
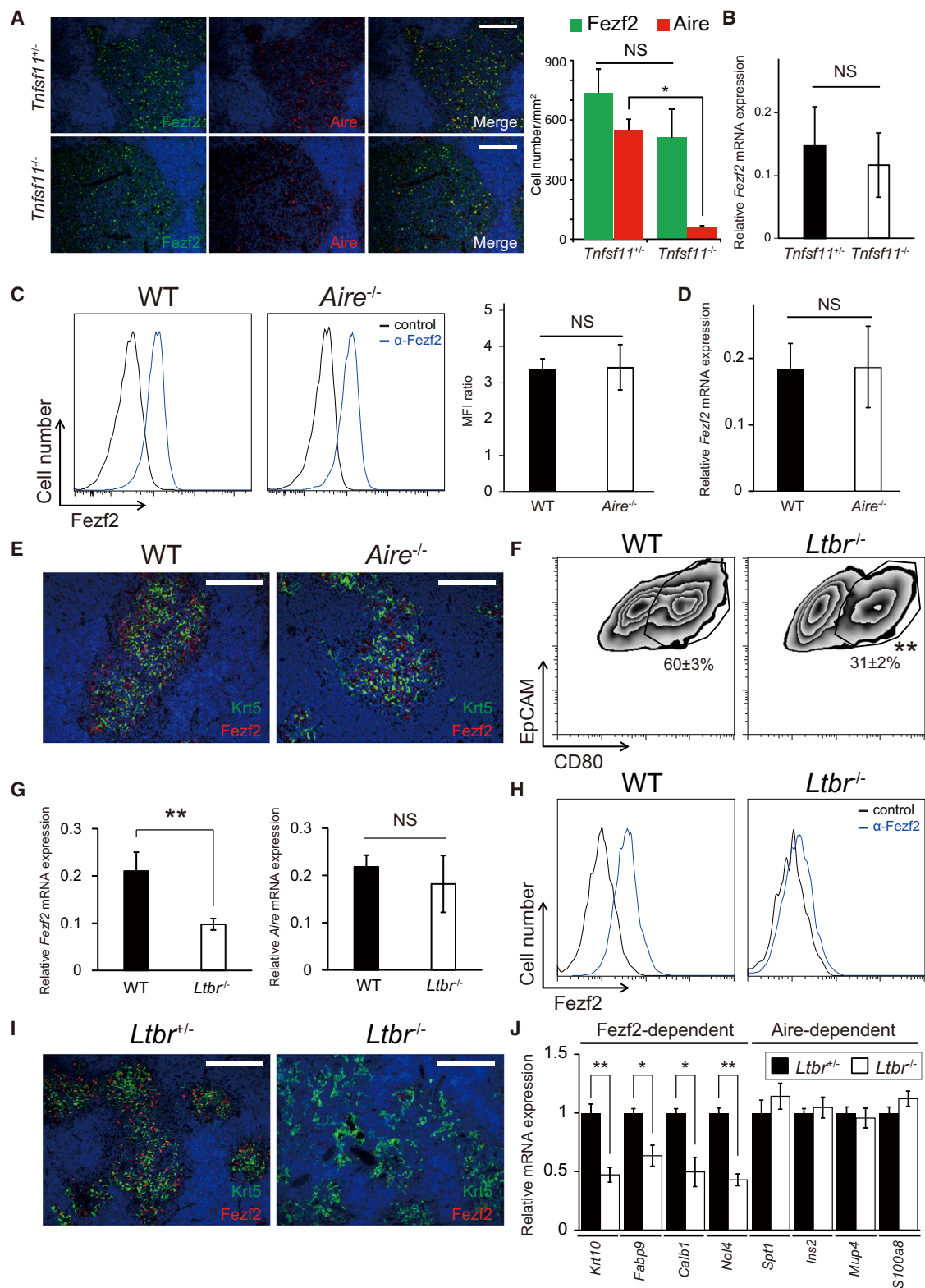
Figure 6. LTβR Pathway Regulates *Fezf2* Expression in mTECs

(A) Fezf2 and Aire protein expression in *Tnfsf11*$^{+/-}$ and *Tnfsf11*$^{-/-}$ thymus (immunostaining). Right panel, the number of Aire$^+$ and Fezf2$^+$ cells in the thymic medulla (n = 3 per genotype).
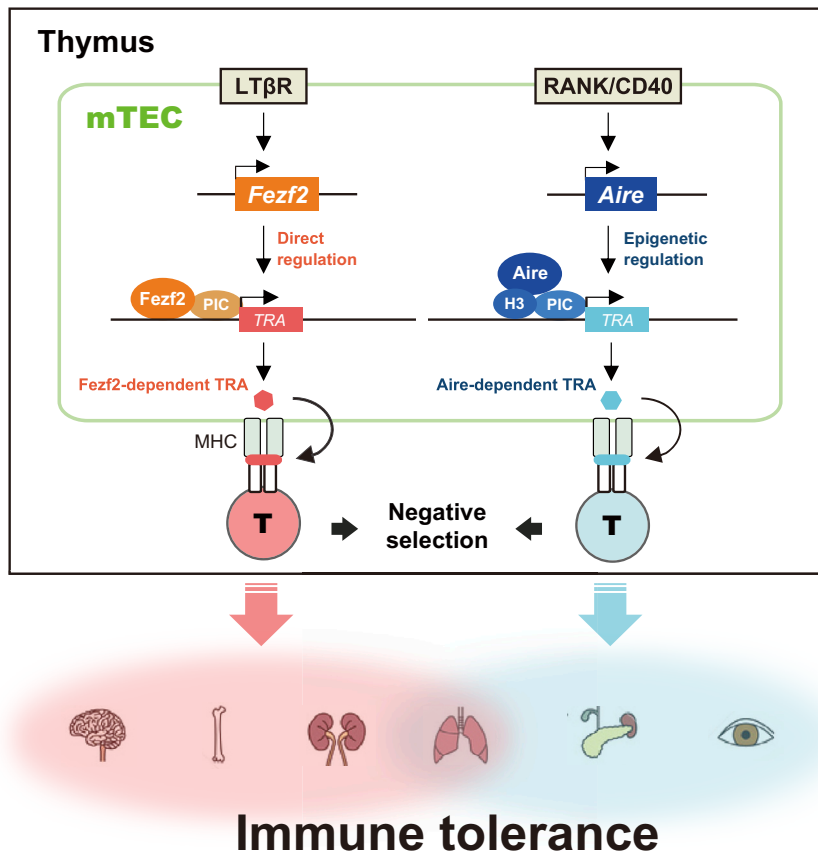
(legend continued on next page)

Figure 7. A Schematic Model of the Thymic Program of TRA Expression Regulated by Fezf2 and Aire

mTEC promiscuously express TRAs, which are dependent on Fezf2 and Aire. The LTβR pathway regulates the expression of Fezf2, which directly induces Fezf2-dependent TRAs by binding to the promoter of these genes. The RANK/CD40 pathways regulate the expression of Aire, which interacts with nuclear proteins (such as histone H3) and controls the expression of Aire-dependent TRAs through epigenetic mechanisms. The LTbR-Fezf2 and RANK/CD40-Aire pathways are both required for the establishment of immune tolerance. PIC, preinitiation complex (including RNA polymerase II and basal transcription factors).

430 2.0 array. Data were analyzed using Gene-Spring and deposited in the GEO database under accession number (GSE69105). Fezf2-dependent genes (GSE69105) and Aire-dependent genes (GSE85) that are common probes were picked up from the platforms (total 8,848 coding genes) and compared (cutoff; 1.5-fold, p value; p < 0.05).

Detection of Inflammatory Cell Infiltration and Autoantibodies
Organs from the mice were harvested and fixed with paraformaldehyde (Nacalai), embedded in paraffin, sectioned, and stained with H&E. Sera from the WT/nu, KO/nu, Aire⁻/⁻, and Foxn1-Cre⁺ Fezf2flox/⁻ mice were prepared from tail vein blood. For the detection of the autoantibodies, serial frozen sections of the lung, joint, eye, and skin from Rag1⁻/⁻ mice were incubated with diluted sera in 0.01% Triton-PBS (1:50) followed by Alexa-Fluor-555-conjugated goat anti-mouse immunoglobulin G (IgG) polyclonal antibodies (Invitrogen).

stromal cells were prepared according to a previous report (Hikosaka et al., 2008), and sorted with a FACSAria III (BD Biosciences).

qRT-PCR
To detect the expression of Fezf2, Aire, and TRA mRNAs, total RNA was extracted from mTECs using an RNeasy mini or micro kit (Invitrogen). Total RNA was subjected to reverse transcription with a first-strand synthesis supermix for qRT-PCR (Invitrogen). The amplification conditions were previously described (Komatsu et al., 2014). The gene expression was normalized by Gapdh expression.

Microarray Analysis
Total RNA of mTECs isolated by cell sorting from WT mice and Fezf2⁻/⁻ mice was extracted with an RNeasy Micro Kit (QIAGEN) and processed for microarray analysis, as previously described (Komatsu et al., 2014). GeneChip analysis was performed using the Affymetrix mouse genome

Luciferase Reporter Assay
Cells from the mTEC line 1C6 were transfected into pNluc plasmids or pNluc-the promoters together with Fezf2 expression or control vectors using Lipofectamine 2000 (Invitrogen) and incubated for 24 hr. Signals were detected with Nano-Glo luciferase assay system (Promega) and quantified with a plate reader (Berthold MicoLumatPlus LB96V 96well Luminometer).

Cytometric Beads Assay
The mouse inflammatory cytokines or chemokines were detected by cytometric bead assay kits (BD Biosciences) and analyzed with a FACSCanto II (BD Biosciences).

(B) qRT-PCR analysis of Fezf2 mRNA expression in mTECs (n = 3 per genotype).
(C) Flow cytometric analysis of Fezf2 protein expression in mTECs from WT and Aire⁻/⁻ mice. Right panel, the mean fluorescence intensity (MFI) ratio (n = 8 per genotype).
(D) Fezf2 mRNA expression in mTECs from WT and Aire⁻/⁻ mice (n = 5 per genotype).
(E) Histochemical analyses of Fezf2 protein expression in the WT and Aire⁻/⁻ thymus. Scale bars, 100 μm.
(F) Flow cytometric analysis of the fraction of mTECs (CD80⁺ cells) among the total TECs (EpCAM⁺CD45⁻ cells) in Ltbr⁺/⁻ and Ltbr⁻/⁻ mice (n = 4 per genotype).
(G) The decreased mRNA expression of Fezf2 in Ltbr⁻/⁻ mTECs compared to the WT (n = 4 per genotype).
(H) Fezf2 protein expression in mTECs in Ltbr⁺/⁻ and Ltbr⁻/⁻ mice (n = 4 per genotype).
(I) Histochemical analyses of Fezf2 protein expression in the Ltbr⁺/⁻ and Ltbr⁻/⁻ thymus. Scale bars, 100 μm.
(J) Fezf2-dependent or Aire-dependent TRA mRNA expression in mTECs in Ltbr⁺/⁻ and Ltbr⁻/⁻ mice (n = 4 per genotype). Statistical data are shown as the mean ± SD. NS, not significant; *p < 0.05; **p < 0.01.
See also Figure S6.

ELISA

The autoantibodies in the mouse serum were analyzed using ELISA kits (Southern Biotech) and quantified with a plate reader (BioRAD iMark Microplate Reader), as previously described (Komatsu et al., 2014). Mouse recombinant proteins were transthyretin (Uscn, RPA726Mu01), monoamine oxidase a (Usbio, 155899), calbindin (Uscn, RPG438Mu01), and interstitial retinol binding protein (Uscn, RPA367Mu01).

Statistical Analysis

Statistical analysis was performed using one-way ANOVA followed by post hoc Bonferroni test when applicable or unpaired two-tailed t test (*p < 0.05; **p < 0.01; ***p < 0.001; NS, not significant, throughout the paper). All data are expressed as the mean ± SD. Results are representative of more than two independent experiments.

Detailed methods can be found in the Supplemental Information.

ACCESSION NUMBERS

Data were analyzed using GeneSpring and deposited in the GEO database under accession number GSE69105.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.013.

AUTHOR CONTRIBUTIONS

H. Takaba designed the study, performed most of the experiments, interpreted the results, and wrote the manuscript. Y.M. prepared the mouse tissue sections and performed the pathological analysis. Y.T. performed biochemical experiments and conducted bioinformatics analysis. L.D., N.K., and T.N. contributed to manuscript preparation and data interpretation. T.K. conducted GeneChip analysis and contributed to data interpretation. H. Takayanagi supervised the project and wrote the manuscript.

ACKNOWLEDGMENTS

REFERENCES

Aaltonen, J., Björses, P., Perheentupa, J., Horelli-Kuitunen, N., Palotie, A., Peltonen, L., Su Lee, Y., Francis, F., Henning, S., Thiel, C., et al.; Finnish-German APECED Consortium (1997). An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. Nat. Genet. *17*, 399–403.

Abramson, J., Giraud, M., Benoist, C., and Mathis, D. (2010). Aire's partners in the molecular control of immunological tolerance. Cell *140*, 123–135.

Akirav, E.M., Ruddle, N.H., and Herold, K.C. (2011). The role of AIRE in human autoimmune disease. Nat. Rev. Endocrinol. *7*, 25–33.

Akiyama, T., Maeda, S., Yamane, S., Ogino, K., Kasai, M., Kajiura, F., Matsumoto, M., and Inoue, J. (2005). Dependence of self-tolerance on TRAF6-directed development of thymic stroma. Science *308*, 248–251.

Akiyama, T., Shimo, Y., Yanai, H., Qin, J., Ohshima, D., Maruyama, Y., Asaumi, Y., Kitazawa, J., Takayanagi, H., Penninger, J.M., et al. (2008). The tumor necrosis factor family receptors RANK and CD40 cooperatively establish the thymic medullary microenvironment and self-tolerance. Immunity *29*, 423–437.

Akiyama, N., Shinzawa, M., Miyauchi, M., Yanai, H., Tateishi, R., Shimo, Y., Ohshima, D., Matsuo, K., Sasaki, I., Hoshino, K., et al. (2014). Limitation of immune tolerance-inducing thymic epithelial cell development by Spi-B-mediated negative feedback regulation. J. Exp. Med. *211*, 2425–2438.

Ali, F., Rowley, M., Jayakrishnan, B., Teuber, S., Gershwin, M.E., and Mackay, I.R. (2011). Stiff-person syndrome (SPS) and anti-GAD-related CNS degenerations: protean additions to the autoimmune central neuropathies. J. Autoimmun. *37*, 79–87.

Anderson, G., and Takahama, Y. (2012). Thymic epithelial cells: working class heroes for T cell development and repertoire selection. Trends Immunol. *33*, 256–263.

Anderson, M.S., Venanzi, E.S., Klein, L., Chen, Z., Berzins, S.P., Turley, S.J., von Boehmer, H., Bronson, R., Dierich, A., Benoist, C., and Mathis, D. (2002). Projection of an immunological self shadow within the thymus by the aire protein. Science *298*, 1395–1401.

Belogurov, A.A., Jr., Kurkova, I.N., Friboulet, A., Thomas, D., Misikov, V.K., Zakharova, M.Y., Suchkov, S.V., Kotov, S.V., Alehin, A.I., Avalle, B., et al. (2008). Recognition and degradation of myelin basic protein peptides by serum autoantibodies: novel biomarker for multiple sclerosis. J. Immunol. *180*, 1258–1267.

Boehm, T., Scheu, S., Pfeffer, K., and Bleul, C.C. (2003). Thymic medullary epithelial cell differentiation, thymocyte emigration, and the control of autoimmunity require lympho-epithelial cross talk via LTbetaR. J. Exp. Med. *198*, 757–769.

Derbinski, J., Schulte, A., Kyewski, B., and Klein, L. (2001). Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. Nat. Immunol. *2*, 1032–1039.

Derbinski, J., Gäbler, J., Brors, B., Tierling, S., Jonnakuty, S., Hergenhahn, M., Peltonen, L., Walter, J., and Kyewski, B. (2005). Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. J. Exp. Med. *202*, 33–45.

Deshmane, S.L., Kremlev, S., Amini, S., and Sawaya, B.E. (2009). Monocyte chemoattractant protein-1 (MCP-1): an overview. J. Interferon Cytokine Res. *29*, 313–326.

DeVoss, J., Hou, Y., Johannes, K., Lu, W., Liou, G.I., Rinn, J., Chang, H., Caspi, R.R., Fong, L., and Anderson, M.S. (2006). Spontaneous autoimmunity prevented by thymic expression of a single self-antigen. J. Exp. Med. *203*, 2727–2735.

Eckler, M.J., and Chen, B. (2014). Fez family transcription factors: controlling neurogenesis and cell fate in the developing mammalian nervous system. BioEssays *36*, 788–797.

Endo, T., Takizawa, S., Tanaka, S., Takahashi, M., Fujii, H., Kamisawa, T., and Kobayashi, T. (2009). Amylase alpha-2A autoantibodies: novel marker of autoimmune pancreatitis and fulminant type 1 diabetes. Diabetes *58*, 732–737.

Fatourou, E.M., and Koskinas, J.S. (2009). Adaptive immunity in hepatocellular carcinoma: prognostic and therapeutic implications. Expert Rev. Anticancer Ther. *9*, 1499–1510.

Gallegos, A.M., and Bevan, M.J. (2004). Central tolerance to tissue-specific antigens mediated by direct and indirect antigen presentation. J. Exp. Med. *200*, 1039–1049.

Giraud, M., Jmari, N., Du, L., Carallis, F., Nieland, T.J., Perez-Campo, F.M., Bensaude, O., Root, D.E., Hacohen, N., Mathis, D., and Benoist, C. (2014). An RNAi screen for Aire cofactors reveals a role for Hnrnpl in polymerase release and Aire-activated ectopic transcription. Proc. Natl. Acad. Sci. USA *111*, 1491–1496.

Gray, D.H., Seach, N., Ueno, T., Milton, M.K., Liston, A., Lew, A.M., Goodnow, C.C., and Boyd, R.L. (2006). Developmental kinetics, turnover, and stimulatory capacity of thymic epithelial cells. Blood 108, 3777–3785.

Han, W., Kwan, K.Y., Shim, S., Lam, M.M., Shin, Y., Xu, X., Zhu, Y., Li, M., and Sestan, N. (2011). TBR1 directly represses Fezf2 to control the laminar origin and development of the corticospinal tract. Proc. Natl. Acad. Sci. USA 108, 3041–3046.

Hikosaka, Y., Nitta, T., Ohigashi, I., Yano, K., Ishimaru, N., Hayashi, Y., Matsumoto, M., Matsuo, K., Penninger, J.M., Takayanagi, H., et al. (2008). The cytokine RANKL produced by positively selected thymocytes fosters medullary thymic epithelial cells that express autoimmune regulator. Immunity 29, 438–450.

Hinterberger, M., Aichinger, M., Prazeres da Costa, O., Voehringer, D., Hoffmann, R., and Klein, L. (2010). Autonomous role of medullary thymic epithelial cells in central CD4(+) T cell tolerance. Nat. Immunol. 11, 512–519.

Hirata, T., Suda, Y., Nakao, K., Narimatsu, M., Hirano, T., and Hibi, M. (2004). Zinc finger gene fez-like functions in the formation of subplate neurons and thalamocortical axons. Dev. Dyn. 230, 546–556.

Hofmann, J., Greter, M., Du Pasquier, L., and Becher, B. (2010). B-cells need a proper house, whereas T-cells are happy in a cave: the dependence of lymphocytes on secondary lymphoid tissues during evolution. Trends Immunol. 31, 144–153.

Hogquist, K.A., Baldwin, T.A., and Jameson, S.C. (2005). Central tolerance: learning self-control in the thymus. Nat. Rev. Immunol. 5, 772–782.

Kappler, J.W., Roehm, N., and Marrack, P. (1987). T cell tolerance by clonal elimination in the thymus. Cell 49, 273–280.

Klein, L., Kyewski, B., Allen, P.M., and Hogquist, K.A. (2014). Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). Nat. Rev. Immunol. 14, 377–391.

Koh, A.S., Kingston, R.E., Benoist, C., and Mathis, D. (2010). Global relevance of Aire binding to hypomethylated lysine-4 of histone-3. Proc. Natl. Acad. Sci. USA 107, 13016–13021.

Komatsu, N., Okamoto, K., Sawa, S., Nakashima, T., Oh-hora, M., Kodama, T., Tanaka, S., Bluestone, J.A., and Takayanagi, H. (2014). Pathogenic conversion of Foxp3+ T cells into TH17 cells in autoimmune arthritis. Nat. Med. 20, 62–68.

Kwan, K.Y. (2013). Transcriptional dysregulation of neocortical circuit assembly in ASD. Int. Rev. Neurobiol. 113, 167–205.

Laan, M., Kisand, K., Kont, V., Möll, K., Tserel, L., Scott, H.S., and Peterson, P. (2009). Autoimmune regulator deficiency results in decreased expression of CCR4 and CCR7 ligands and in delayed migration of CD4+ thymocytes. J. Immunol. 183, 7682–7691.

Lei, Y., Ripen, A.M., Ishimaru, N., Ohigashi, I., Nagasawa, T., Jeker, L.T., Bösl, M.R., Holländer, G.A., Hayashi, Y., Malefyt, Rde.W., et al. (2011). Aire-dependent production of XCL1 mediates medullary accumulation of thymic dendritic cells and contributes to regulatory T cell development. J. Exp. Med. 208, 383–394.

Liston, A., Lesage, S., Wilson, J., Peltonen, L., and Goodnow, C.C. (2003). Aire regulates negative selection of organ-specific T cells. Nat. Immunol. 4, 350–354.

Locksley, R.M., Killeen, N., and Lenardo, M.J. (2001). The TNF and TNF receptor superfamilies: integrating mammalian biology. Cell 104, 487–501.

Lodato, S., Molyneaux, B.J., Zuccaro, E., Goff, L.A., Chen, H.H., Yuan, W., Meleski, A., Takahashi, E., Mahony, S., Rinn, J.L., et al. (2014). Gene co-regulation by Fezf2 selects neurotransmitter identity and connectivity of corticospinal neurons. Nat. Neurosci. 17, 1046–1054.

Martins, V.C., Boehm, T., and Bleul, C.C. (2008). Ltbetar signaling does not regulate Aire-dependent transcripts in medullary thymic epithelial cells. J. Immunol. 181, 400–407.

Mathis, D., and Benoist, C. (2009). Aire. Annu. Rev. Immunol. 27, 287–312.

Mouri, Y., Yano, M., Shinzawa, M., Shimo, Y., Hirota, F., Nishikawa, Y., Nii, T., Kiyonari, H., Abe, T., Uehara, H., et al. (2011). Lymphotoxin signal promotes thymic organogenesis by eliciting RANK expression in the embryonic thymic stroma. J. Immunol. 186, 5047–5057.

Nagamine, K., Peterson, P., Scott, H.S., Kudoh, J., Minoshima, S., Heino, M., Krohn, K.J., Lalioti, M.D., Mullis, P.E., Antonarakis, S.E., et al. (1997). Positional cloning of the APECED gene. Nat. Genet. 17, 393–398.

Nishikawa, Y., Hirota, F., Yano, M., Kitajima, H., Miyazaki, J., Kawamoto, H., Mouri, Y., and Matsumoto, M. (2010). Biphasic Aire expression in early embryos and in medullary thymic epithelial cells before end-stage terminal differentiation. J. Exp. Med. 207, 963–971.

Peterson, P., Org, T., and Rebane, A. (2008). Transcriptional regulation by AIRE: molecular mechanisms of central tolerance. Nat. Rev. Immunol. 8, 948–957.

Roulois, D., Grégoire, M., and Fonteneau, J.F. (2013). MUC1-specific cytotoxic T lymphocytes in cancer therapy: induction and challenge. BioMed Res. Int. 2013, 871936.

Sakaguchi, S., Ono, M., Setoguchi, R., Yagi, H., Hori, S., Fehervari, Z., Shimizu, J., Takahashi, T., and Nomura, T. (2006). Foxp3+ CD25+ CD4+ natural regulatory T cells in dominant self-tolerance and autoimmune disease. Immunol. Rev. 212, 8–27.

Saltis, M., Criscitiello, M.F., Ohta, Y., Keefe, M., Trede, N.S., Goitsuka, R., and Flajnik, M.F. (2008). Evolutionarily conserved and divergent regions of the autoimmune regulator (Aire) gene: a comparative analysis. Immunogenetics 60, 105–114.

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485, 237–241.

Sansom, S.N., Shikama-Dorn, N., Zhanybekova, S., Nusspaumer, G., Macaulay, I.C., Deadman, M.E., Heger, A., Ponting, C.P., and Holländer, G.A. (2014). Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. Genome Res. 24, 1918–1931.

Seach, N., Ueno, T., Fletcher, A.L., Lowen, T., Mattesich, M., Engwerda, C.R., Scott, H.S., Ware, C.F., Chidgey, A.P., Gray, D.H., and Boyd, R.L. (2008). The lymphotoxin pathway regulates Aire-independent expression of ectopic genes and chemokines in thymic stromal cells. J. Immunol. 180, 5384–5392.

Sharma, S., Ghosh, S., Singh, L.K., Sarkar, A., Malhotra, R., Garg, O.P., Singh, Y., Sharma, R.S., Bhakuni, D.S., Das, T.K., and Biswas, S. (2014). Identification of autoantibodies against transthyretin for the screening and diagnosis of rheumatoid arthritis. PLoS ONE 9, e93905.

Shim, S., Kwan, K.Y., Li, M., Lefebvre, V., and Sestan, N. (2012). Cis-regulatory control of corticospinal system development and evolution. Nature 486, 74–79.

Shu, X.S., Li, L., Ji, M., Cheng, Y., Ying, J., Fan, Y., Zhong, L., Liu, X., Tsao, S.W., Chan, A.T., and Tao, Q. (2013). FEZF2, a novel 3p14 tumor suppressor gene, represses oncogene EZH2 and MDM2 expression and is frequently methylated in nasopharyngeal carcinoma. Carcinogenesis 34, 1984–1993.

Takaba, H., Imai, T., Miki, S., Morishita, Y., Miyashita, A., Ishikawa, N., Nishizumi, H., and Sakano, H. (2013). A major allogenic leukocyte antigen in the agnathan hagfish. Sci. Rep. 3, 1716.

Venanzi, E.S., Gray, D.H., Benoist, C., and Mathis, D. (2007). Lymphotoxin pathway and Aire influences on thymic medullary epithelial cells are unconnected. J. Immunol. 179, 5693–5700.

Waterfield, M., Khan, I.S., Cortez, J.T., Fan, U., Metzger, T., Greer, A., Fasano, K., Martinez-Llordella, M., Pollack, J.L., Erle, D.J., et al. (2014). The transcriptional regulator Aire coopts the repressive ATF7ip-MBD1 complex for the induction of immunotolerance. Nat. Immunol. 15, 258–265.

White, A.J., Nakamura, K., Jenkinson, W.E., Saini, M., Sinclair, C., Seddon, B., Narendran, P., Pfeffer, K., Nitta, T., Takahama, Y., et al. (2010). Lymphotoxin signals from positively selected thymocytes regulate the terminal differentiation of medullary thymic epithelial cells. J. Immunol. 185, 4769–4776.

Yang, S., Fujikado, N., Kolodin, D., Benoist, C., and Mathis, D. (2015). Immune tolerance. Regulatory T cells generated early in life play a distinct role in maintaining self-tolerance. Science 348, 589–594.

Cell

Dissecting Polyclonal Vaccine-Induced Humoral Immunity against HIV Using Systems Serology

Graphical Abstract

Highlights

- Beyond neutralization, antibodies drive antiviral control via Fc-mediated functions

- Distinct vaccines elicit unique antibody Fc-effector profiles

- Network analyses comprehensively integrating antibody profiles can compare vaccines

- Case:control RV144 analysis points to mechanisms of reduced risk of HIV infection
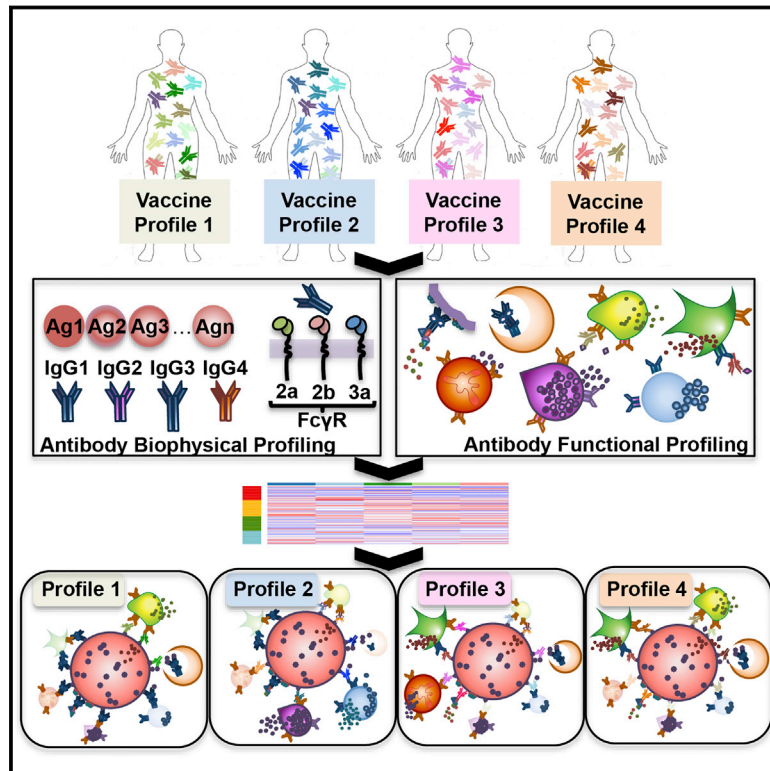
Authors

Amy W. Chung, Manu P. Kumar, Kelly B. Arnold, ..., Dan H. Barouch, Douglas A. Lauffenburger, Galit Alter

Correspondence

lauffen@mit.edu (D.A.L.), galter@mgh.harvard.edu (G.A.)

In Brief

Systems Serology reveals unique vaccine-induced "fingerprints," highlighting potential markers of protection against HIV and providing a powerful method for comparing candidate vaccines against pathogens for which correlates of protection remain elusive.

CellPress

Dissecting Polyclonal Vaccine-Induced Humoral Immunity against HIV Using Systems Serology

Amy W. Chung,[1,2,12] Manu P. Kumar,[3,12] Kelly B. Arnold,[3,12] Wen Han Yu,[1,3,12] Matthew K. Schoen,[1] Laura J. Dunphy,[3] Todd J. Suscovich,[1] Nicole Frahm,[4] Caitlyn Linde,[1] Alison E. Mahan,[1] Michelle Hoffner,[1] Hendrik Streeck,[5,6] Margaret E. Ackerman,[7] M. Juliana McElrath,[4] Hanneke Schuitemaker,[8] Maria G. Pau,[8] Lindsey R. Baden,[1,9] Jerome H. Kim,[5,10] Nelson L. Michael,[5] Dan H. Barouch,[1,11] Douglas A. Lauffenburger,[3,*] and Galit Alter[1,*]

[1]Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA 02139, USA
[2]Department of Microbiology and Immunology, Peter Doherty Institute, University of Melbourne, Parkville, VIC 3010, Australia
[3]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[5]Department of Molecular Virology and Pathogenesis, Walter Reed Army Institute of Research, U.S. Military HIV Research Program, Silver Spring, MD 20910, USA
[6]Institute for Medical Biology, University Hospital Essen, University Duisburg-Essen, Essen 45141, Germany
[7]Thayer School of Engineering at Dartmouth, Hanover, NH 03755, USA
[8]Crucell Holland B.V., The Janssen Pharmaceutical Companies of Johnson & Johnson, Leiden 2333, the Netherlands
[9]Division of Infectious Diseases, Brigham and Women's Hospital, Boston, MA 02215, USA
[10]International Vaccine Institute, Seoul 151-742, Republic of Korea
[11]Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA
[12]Co-first author
*Correspondence: lauffen@mit.edu (D.A.L.), galter@mgh.harvard.edu (G.A.)
http://dx.doi.org/10.1016/j.cell.2015.10.027

SUMMARY

While antibody titers and neutralization are considered the gold standard for the selection of successful vaccines, these parameters are often inadequate predictors of protective immunity. As antibodies mediate an array of extra-neutralizing Fc functions, when neutralization fails to predict protection, investigating Fc-mediated activity may help identify immunological correlates and mechanism(s) of humoral protection. Here, we used an integrative approach termed Systems Serology to analyze relationships among humoral responses elicited in four HIV vaccine trials. Each vaccine regimen induced a unique humoral "Fc fingerprint." Moreover, analysis of case:control data from the first moderately protective HIV vaccine trial, RV144, pointed to mechanistic insights into immune complex composition that may underlie protective immunity to HIV. Thus, multi-dimensional relational comparisons of vaccine humoral fingerprints offer a unique approach for the evaluation and design of novel vaccines against pathogens for which correlates of protection remain elusive.

INTRODUCTION

Although over 80 vaccines, covering more than 20 diseases, have been licensed in the United States, vaccine design efforts against persisting infections, including malaria, tuberculosis, and HIV, continue to fail. These setbacks have driven a shift from empirical vaccine design approaches to rational vaccine development strategies that consider pathogen life cycle, pathogen structural information, and immunological correlates of protection. However, the immune correlates for most globally lethal pathogens have yet to be defined, complicating vaccine design efforts. Prospective immunogens are frequently chosen based on measures of antibody (Ab) titer and neutralization, regardless of their mechanistic effects in immunity. However, for most clinically approved vaccines, titer and neutralization activity alone do not account for protective immunity (Pulendran and Ahmed, 2011). Instead, protective immunity is often observable in the absence of neutralization, and accumulating evidence across a spectrum of vaccines has suggested a critical role for extra-neutralizing Ab functions such as Ab-dependent cellular cytotoxicity (ADCC), Ab-dependent cellular phagocytosis (ADCP), Ab-dependent complement deposition (ADCD), and Ab-dependent respiratory burst (ADRB) in both protection from and post-infection control of HIV (Barouch et al., 2015; Bournazos et al., 2014; Hessell et al., 2007), influenza (DiLillo et al., 2014; Jegerlehner et al., 2004), herpes simplex virus (HSV) (Kohl and Loo, 1982; Kohl et al., 1981), Ebola virus (Warfield et al., 2007), and malaria (Joos et al., 2010; Osier et al., 2014).

Following vaccination, Abs targeting an extensive array of epitopes with different affinities and Fc-effector profiles collectively contribute to the formation of immune complexes that direct antimicrobial functions via their constant domains (Fc). In addition to the rapid diversification of the antigen (Ag)-binding domain (Fab), the Fc domain is also rapidly tuned during an immune response, altering the affinity of Ab interactions with innate immune receptors (e.g., Fc receptors and complement) expressed on all innate immune cells (Ackerman and Alter, 2013; Chung and Alter, 2014). The diversity of Fc profiles, potential Fab variants, and tissue-specific Fc receptor expression results
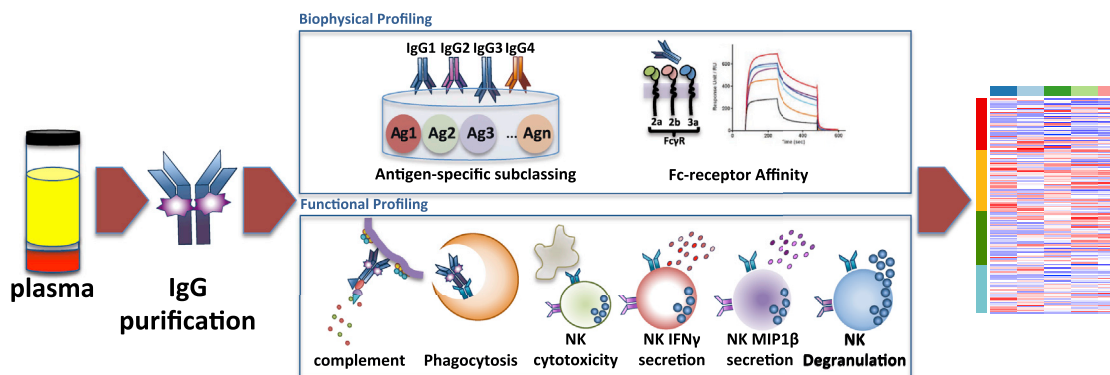
Figure 1. System Serology Analysis
This Systems Serology platform allows for the broad characterization of the polyclonal extra-neutralizing IgG immune profile induced by vaccination. IgG was purified from subjects enrolled in four different HIV vaccine trials (RV144, VAX003, HVTN204, and IPCAVD001). Six Fc-effector functions and 58 biophysical measurements were assayed (complete list described in Table S1). All 64 parameters were collected to create an extra-neutralizing serological signature for the four vaccine trials, using an array of unsupervised and supervised machine learning algorithms.
See also Tables S1 and S2.

in a flexible humoral immune response poised for the elimination of pathogens via mechanisms beyond simple neutralization. Hence, analytical approaches able to integrate diverse facets of the humoral immune response will be critical to: (1) define unexpected correlates of protection from infection in protection studies or studies of natural disease resistance, (2) guide the selection of promising vaccines/immunogens through principled analysis of humoral immune profiles, and (3) define the relationships between Ab populations and functions that point to mechanisms of protective immunity.

As a prominent example, the ability to select HIV vaccine candidates has been hindered by an inadequate understanding of the immunological correlates of protection from HIV. However, several clinical trials have been conducted, one of which (RV144) demonstrated a modest level of protection (31.2% reduction in the risk of infection) (Rerks-Ngarm et al., 2009), potentially harboring clues that may guide future vaccine development. This protection was observed in the absence of neutralizing Abs, cytotoxic T-cell responses, and high Ab titers. Univariate and multivariate logistic regression analyses linked the reduced risk of infection with non-immunoglobulin (Ig)A Ab responses targeting the V1V2 region of the HIV envelope and ADCC activity (Haynes et al., 2012; Zolla-Pazner et al., 2014). Follow-up analyses identified additional features of the humoral immune response associated with protection, including the preferential induction of IgG3 responses, which coordinated multiple Ab effector functions, including ADCC and ADCP (Chung et al., 2014b; Yates et al., 2014). However, in the correlates analysis, although many Ab assays were initially considered, the identification of immune correlates in RV144 was constrained by the selected assays that deeply interrogated neutralization and Ab specificity but profiled only a limited set of Fc features, including only a few Ab subclasses/isotypes (IgG, IgG3, IgA) and a single function, ADCC.

Here, we aimed to consider more integrative and network-oriented relationships between a broader array of polyclonal Ab features and functional properties associated with vaccine regimens and outcomes. As an initial test of this approach, termed "Systems Serology," we examined recent HIV vaccine trials, including that of the moderately protective RV144 vaccine ALVAC/AIDSVAX B/E (Rerks-Ngarm et al., 2009), two trials that did not demonstrate efficacy in phase 2b trials, (VAX003; AIDSVAX B/E [Pitisuttithum et al., 2006] and HVTN204; DNA/rAD5 [Churchyard et al., 2011]), and one experimental phase 1 study designed to evaluate the prototype vaccine Ad26 vector (IPCAVD001; Ad26.ENVA.01) (Barouch et al., 2013a). A battery of modeling techniques that emphasize co-variation among measurements was applied to these data, revealing features of vaccine–induced "fingerprints" that offer new insights concerning polyclonal Ab immune responses elicited by vaccines.

RESULTS

Systems Serology
Beyond their role in neutralization, Abs mediate a vast array of additional functions via their Fc domains. Thus, a Systems Serology approach was developed to broadly profile the extra-neutralizing Ab activity of vaccine-induced polyclonal Abs (Figure 1). The initial platform interrogated six Fc-effector functions (ADCC, ADCP, ADCD, and three Ab-dependent natural killer (NK) cell activities (Figure 1). Linked to these six functions, 58 biophysical measurements were simultaneously captured, including binding to Fcγ receptors (FCGRs) and the relative abundances of an array of Ag-specific Abs (Table S1) in 120 samples from four HIV vaccine trials (see Supplemental Experimental Procedures).

Identification of Vaccine-Specific Signatures
Unsupervised hierarchical clustering grouped vaccine regimens primarily by immunogen type (Figure 2; Table S2), including an adenovirus (Ad) vector cluster composed of mixed HVTN204 (DNA/Ad5) and IPCAVD001 (Ad26) samples (Figure 2, cluster 1: green and yellow, respectively) and a protein immunogen cluster containing largely mixed VAX003 (protein alone) and RV144
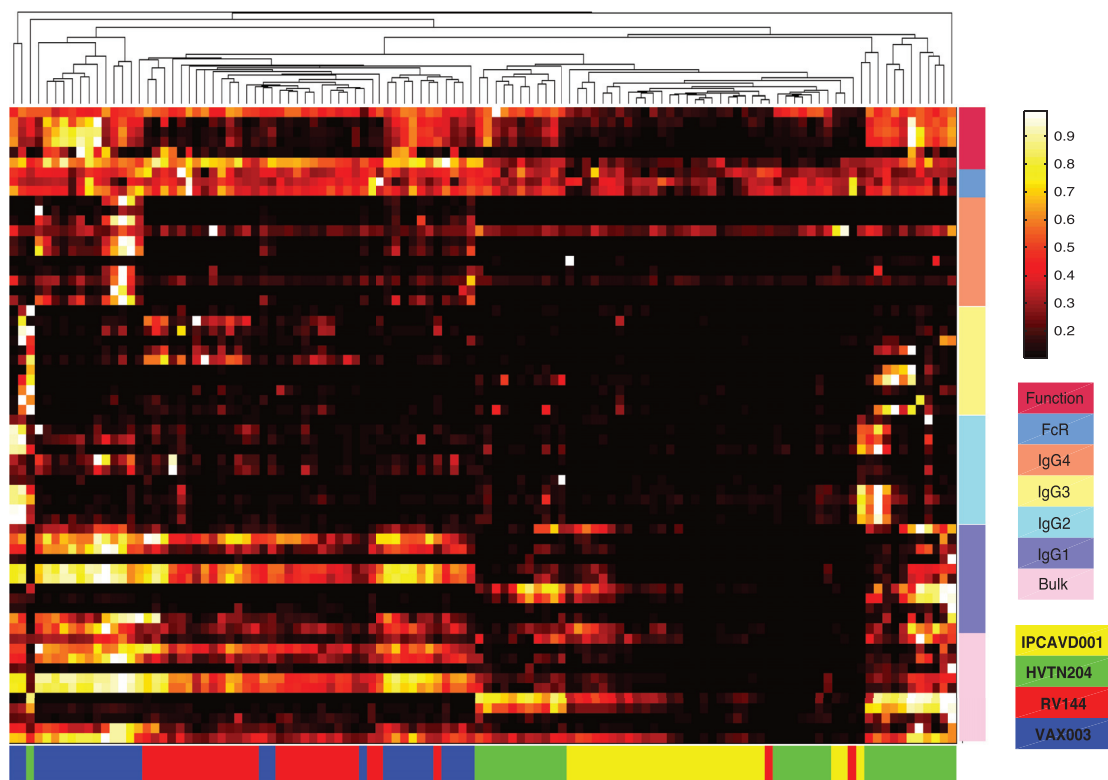
Figure 2. Hierarchical Clustering of Vaccine Trial Profiles by Biophysical Properties and Functional Responses
Data were compiled for the four different vaccine trials. Each column represents the full Ab profile of an individual subject. Colored bars along the bottom correspond to the vaccine trial for each subject. Ab properties are grouped by generalized features (Function, FcR affinity, Bulk IgG, IgG1, IgG2, IgG3, and IgG4), indicated by the colored bars on the right. Specific features are listed in Table S2.
See also Table S1.

(poxvirus prime/protein boost) samples (Figure 2, cluster 2: blue and red, respectively). While this clustering highlights the dominant influence of immunogen type in directing distinct humoral profiles, specific features driving this separation cannot be clearly discerned.

To gain enhanced resolution on the key features contributing to profile differences, a multidimensional combined feature selection method (the least absolute shrinkage and selection operator; LASSO) (Tibshirani, 1997) and partial least-squares discriminant analysis (PLSDA) (Arnold et al., 2015; Lau et al., 2011) were used. Focusing initially on RV144 and VAX003, which shared the same protein immunogen but provided different efficacies, as few as 7 of the 64 features accounted for 76% of the variance across the two trials, driving nearly complete resolution of the vaccine profiles (Figures 3A and 3B). Separation of the Ab profiles was observed in the scores plot, with points representing individual RV144 (red) or VAX003 (blue) vaccinees (Figure 3A). Differences between vaccine-elicited Ab profiles were largely captured along the first dimension (LV1), which accounted for the majority of the variance between the two trials (61%). The corresponding loadings plot (Figure 3B) illustrates the contribution of the seven LASSO features, where the relative location of an individual feature is associated with the corresponding vaccine subpopulation in the scores plot (Figure 3A). Elevated

gp120-specific IgG3 levels, relative to other features (Figure 3B), uniquely marked the RV144 vaccine profile (Figure 3A), as previously described (Chung et al., 2014b; Yates et al., 2014). By contrast, the VAX003 Ab profile was associated with known induction of higher total HIV gp140-specific Ab titers, dominated by IgG4 (Chung et al., 2014b). However, additional novel features were identified that associated with the non-protective VAX003 profile, including elevated total gp140-specific responses, higher Ab-driven NK cell degranulation, and chemokine secretion. This result suggests that differences in relationships between Ab features rather than the total Ab amount may be essential for resolving "protective" from "non-protective" vaccine profiles. Moreover, the scores plot highlights an unappreciated level of heterogeneity among the RV144 vaccinees, with respect to the magnitude of the IgG3 response, where 26% of the RV144 vaccinees exhibited a more highly skewed IgG3 response specifically across the second dimension, LV2 (Figure 3A).

When all four vaccine trials were analyzed simultaneously, 15 of the 64 features separated the vaccine profiles, accounting for 57% of the variance. The first dimension (LV1) revealed a similar separation as the hierarchical clustering analysis, separating based on protein (Figure 3C, right) versus Ad-based vectored immunization (33% variance), confirming the dominant effect of immunogen type in directing humoral profiles. LV1
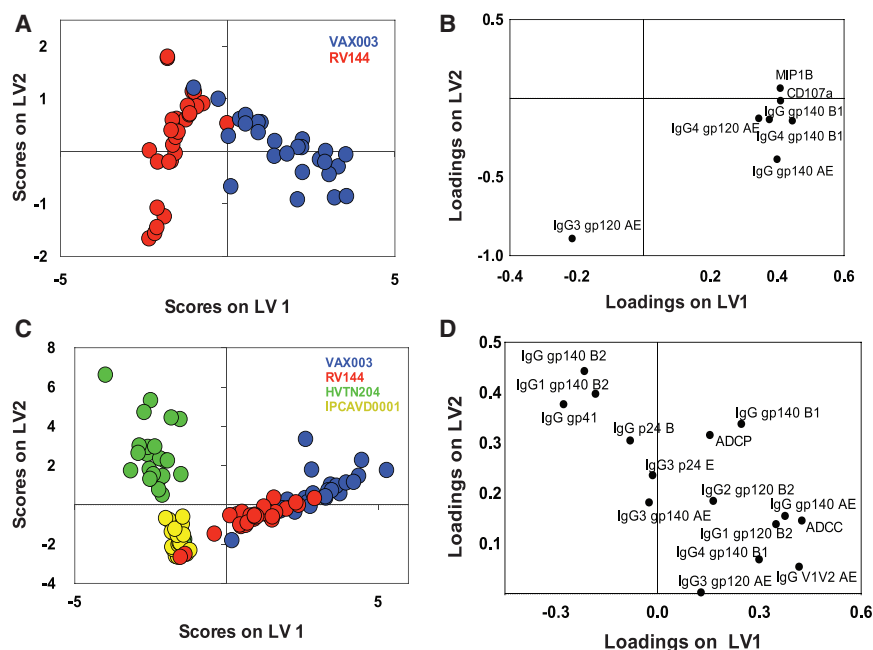
separation was strongly driven by gp41-skewed immunity, due to gp41 being included in the Ad regimens but not included in VAX003 and only partially included in RV144. Furthermore, Abs targeting clade AE Ags (gp120 and V1V2) uniquely marked RV144 and VAX003 profiles, as subjects were immunized with clade AE-derived immunogens. Thus the Ag itself, rather than the vector/immunization regimen alone, was a critical determinant influencing vaccine-induced humoral profiles.

The second dimension identified additional features that further split the vaccine profiles, accounting for an additional 24% of the variance, contributing to an unexpected grouping and separation of RV144/Ad26 and VAX003/Ad5 profiles. This separation was primarily related to differences in IgG3 subclass and V1V2 levels, which scattered in multidimensional space more closely with RV144 and Ad26 profiles (Figure 3D). Therefore, markers previously associated with reduced risk of infection in RV144 co-segregated with the experimental Ad26 vaccine trial, which used a vector similar to the ones used in regimens recently shown to protect non-human primates from infection through non-neutralizing polyfunctional Abs (Barouch et al., 2015).

Thus, use of the LASSO and PLSDA, incorporating co-variation between features, identified key variables involved in classifying vaccine regimens and provided enhanced resolution of the specific Ab features associated with differentiating vaccine profiles, objectively identifying novel correlates of Ab-mediated protection.

Correlation Networks Highlight Distinct Humoral Relationships

Next, we aimed to gain insights into relationships between features contributing to differences among vaccine-induced polyclonal profiles, adapting correlation network analysis tools commonly used in the transcriptomics field. The resulting network models revealed remarkably different Ab co-regulation interactions among the vaccine regimens, providing novel insights into the specific Ab features that may contribute to unique vaccine effector profiles.

VAX003 exhibited the most interconnected network, comprising four dense subnetworks (Figure 4A). The most prominent subnetworks included an unusual tightly tethered mixture of IgG2 and IgG3 responses that are rarely co-selected (Chaudhuri and Alt, 2004; Chaudhuri et al., 2007), pointing to the induction of a non-coherent poorly coordinated functional response (Chung et al., 2014b). Interestingly, all Fc-effector functions were connected to a third subnetwork consisting largely of IgG1 and bulk IgG responses specific for a broad array of Ags that was unexpectedly connected to the fourth IgG4 subnetwork. IgG4 Abs have previously shown to compete actively for immune complex occupancy, resulting in dampened Ab function (Chung et al., 2014b). Thus, the VAX003 network exhibited linked IgG1/IgG4 responses staggered next to a dense IgG2/IgG3 cluster, highlighting the peculiar subclass co-selection profiles driven by the non-protective VAX003 strategy.

While less prominent clusters emerged in the RV144 network model (Figure 4B), ADCP, ADCD, and ADCC were largely tethered to a network of gp120-, gp140- or V1V2-specific IgG1 and/or IgG3 responses. The V1V2B-specific IgG3 response was highly associated with the large IgG1 network, suggesting that high IgG3 V1V2B-specific responses act as a critical surrogate of a coordinated IgG3 and IgG1 response. The total IgG V1V2AE response was directly tethered to both ADCP and ADCC, suggesting that this specific V1V2 response may play an influential role in driving Ab functionality. IgG3 V1V2B and IgG3 V1V2AE responses were not directly correlated, suggesting that these V1V2 responses may represent disparate humoral immune responses rather than a single cross-reactive response. Moreover, because depletion of IgG3 only results in 30% reduction in Ab functionality (Chung et al., 2014b), it is likely that IgG3 responses may serve as a surrogate for a subpopulation
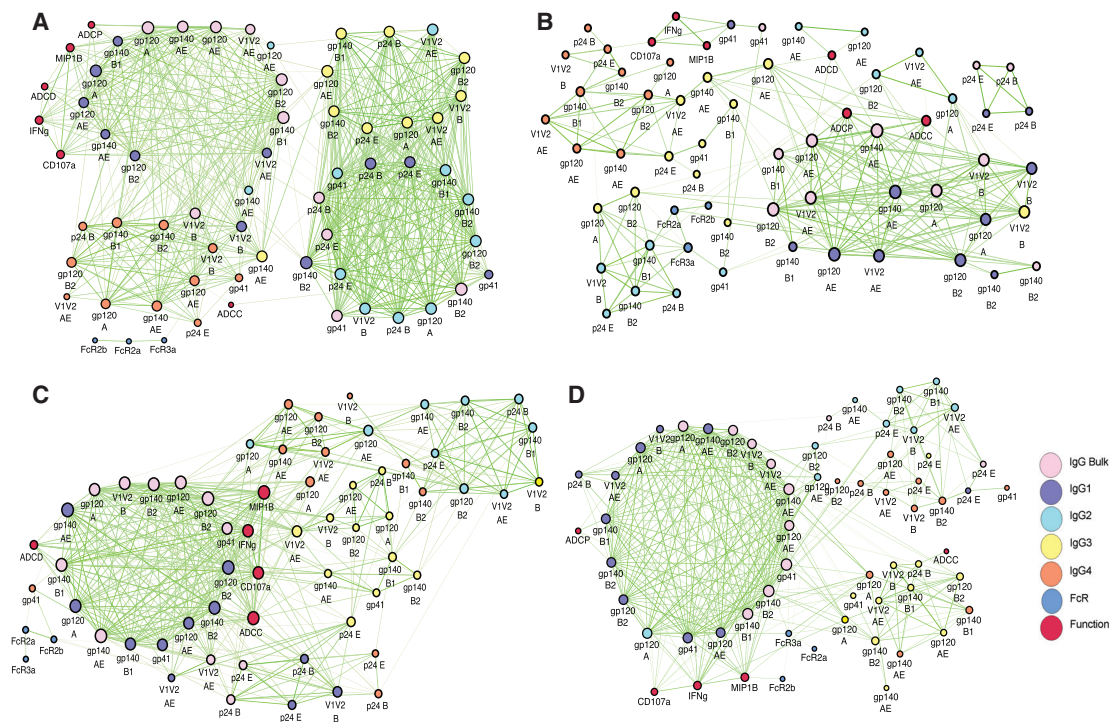
Figure 4. Correlation Networks of Vaccine-Trial-Elicited Humoral Immune Responses Probe Immune Complex Dynamics

(A–D) Correlation networks were generated for VAX003 (A), RV144 (B), HVTN204 (C), and IPCAVD001 (D). Each node (circle) represents either a biophysical feature or an effector function. Nodes are connected with an edge (line) if they are significantly correlated. The different Ab isotypes are identified by different colors as indicated. Edge thickness and color intensity of the connecting lines are directly proportional to statistical significance and edge weight, respectively (thicker and brighter network interactions represent a stronger correlation). The size of each node is directly proportional to its degree of connectedness (i.e., the number of features to which that node is connected).

See also Figure S1 and Tables S1 and S2.

of vaccine-induced IgG1 Abs that direct the polyfunctional Ab responses observed in RV144.

The HVTN204 (DNA/Ad5) network (Figure 4C) contained a highly connected subnetwork, with multiple tethers to less well-connected subnetworks of additional Ab subclasses. The dominant subnetwork consisted of IgG1 Env- and V1V2-specific responses, with ADCC, ADCD, and NK cell responses tightly intercalated within the subnetwork, sandwiched between IgG1 and IgG3 responses. However, ADCP did not appear in the network. This exclusion of ADCP suggests that Ad5 and/or DNA may preclude the induction of phagocytic Ab responses, which have been linked to protection from SIV acquisition (Barouch et al., 2013b).

Conversely, the vaccine profile induced by the experimental IPCAVD001 (Ad26) exhibited a nearly single, densely connected network tethered to Ab functions and Fc-receptor binding activity (Figure 4D). The large network consisted of a tight grid of related bulk IgG/IgG1 responses, while IgG2, IgG3, and IgG4 formed sparse external clusters, including a less functional, interconnected IgG2/IgG4 cluster (Figure 4D, top right). The clear linkages between Ab functions and IgG1 features, including an IgG1 V1V2-driven ADCP response, further supports the potential role of IgG3 as a surrogate of a highly effective, polyfunctional IgG1 response.

Overall, these statistically robust network analyses (Figure S1) point to unique relationships between all features and functions among the four vaccine trials. Identification of "desirable" Ab networks delineating specific biophysical Ab feature/function relationships that are associated with protective immunity may help identify mechanisms underlying correlates, such as the association of IgG3 and V1V2 features with reduced risk of HIV infection by RV144.

System Serology Analysis of Interactions between RV144 Surrogates of Reduced Risk of Infection

Systems Serology approaches can complement existing methods for identifying predictive mechanism(s) of protective immunity. While logistic regression involves stepwise evaluation of strongly correlated individual variables, Systems Serology approaches can additionally identify relationships between Ab features that are predictive of protection. Toward this purpose, we next examined RV144 profiles segregating with known correlates of reduced risk of infection. Thus, we dissected two Ab features (IgG/IgG3 V1V2), which were previously positively associated with reduced acquisition in the RV144 case:control analysis, in our cohort of uninfected vaccinees. Importantly, IgA levels were also included in this analysis, due to their implicated role as correlates of risk (Haynes et al., 2012). Profiles were then compared between "responders" (top 33% for each
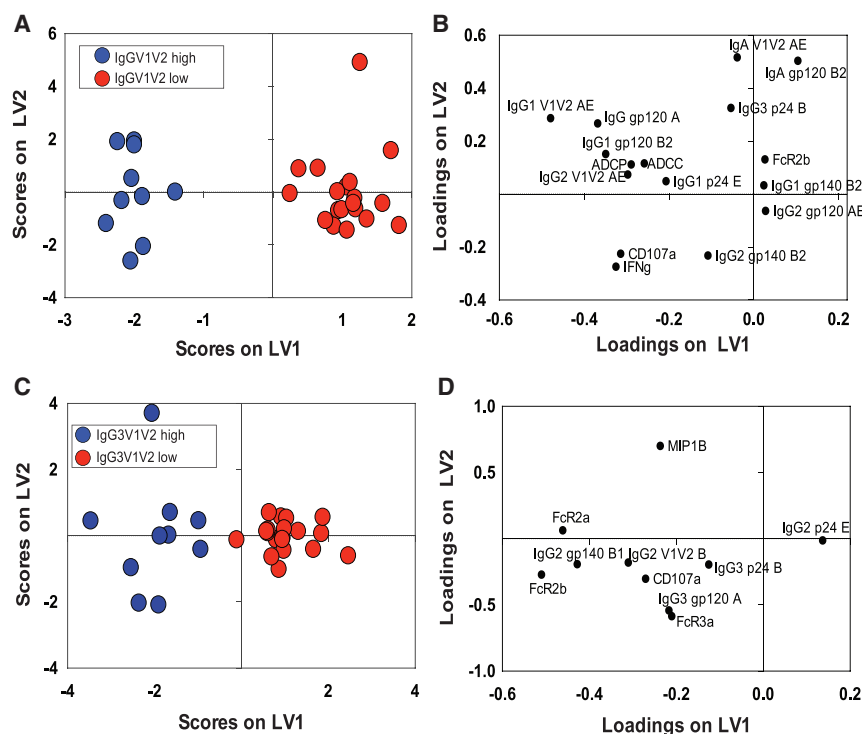
correlate of reduced risk) (Rerks-Ngarm et al., 2009; Zolla-Pazner et al., 2014) or "non-responders." The IgG V1V2 responder profile (Zolla-Pazner et al., 2014) was driven by 16 features (Figures 5A and 5B), including elevated V1V2 responses and a polyfunctional Fc-effector profile linked to higher Ab-dependent NK cell degranulation (i.e., CD107a and interferon γ [IFNγ] expression), ADCP, and ADCC. Conversely, the non-responders exhibited elevated gp120-specific IgA and increased binding to FCGR2B, the sole inhibitory Fcγ receptor, both features that have been previously associated with antagonism of Fc-effector activity (Tomaras et al., 2013; White et al., 2014).

Similar analysis of the IgG3 V1V2 correlate of reduced risk pointed to ten Ab features that distinguished responder/non-responder profiles (Figures 5C and 5D) marked by increased broad Fcγ receptor binding among responders—particularly to activating FCGR2A, involved in ADCP, and to FCGR3A, critical for NK cell degranulation and chemokine secretion. Surprisingly, IgA was not selected as a negative predictor of the IgG3 V1V2 responder profile. These findings confirm that IgG V1V2 (Figure 5A) responders exhibit a balanced polyfunctional profile, while IgG3 V1V2 responders (Figure 5B) possessed Abs selectively enhanced for binding to FCGR2A, associated with ADCP, that has been linked to protection in nonhuman primates (NHPs) (Barouch et al., 2013b).

Defining Integrative Signatures of Protective Humoral Immune Profiles in RV144

Finally, to assess whether our approach could provide enhanced resolution of mechanism(s) of potential reduced risk of infection in the RV144 trial, we next analyzed data from the case:control study (Haynes et al., 2012). Specifically, data characterizing distinct Ab subclass levels targeting multiple vaccine Ags and

functions comparable to those included in our original profiling data were included in the analysis. PLSDA using data from all cases and controls separated placebos from vaccinees, as expected, along LV1 (Figure 6A). In contrast, PLSDA of vaccinees alone was unable to separate the 40 infected from the 201 uninfected vaccinees included in the case:control analysis (Figures S2A and S2B). Similarly, network analyses showed only modest differences between vaccinated cases and controls (Figures S2C and S2D), likely related to the fact that it is unclear which uninfected vaccinees were actually exposed and protected.

To address this complication, we defined groups representing extreme profiles based on known correlates of risk (Haynes et al., 2012). Given that the IgG3 and IgG V1V2 levels were highly correlated (Figure S3), we elected to focus on the IgG V1V2 and IgA relationship due to the intriguing relationships found for these two parameters in the non-case:control data (Figure 5A). Two sets of samples were identified: (1) a region containing the greatest ratio of uninfected:infected vaccinees was classified as the "low-risk" group (Figure 6C, blue box; percentage difference = 28%, p = 0.0088), and (2) the area that contained the lowest ratio of uninfected:infected vaccinees was classified as the "risk" group (Figure 6C, red box; percentage difference = −26%, p = 0.0003). As expected, the lowest frequency of infections was observed in the IgG V1V2^{high}/IgA^{low} region of the plot, and the highest frequency of cases was observed in the IgG V1V2^{low}/IgA^{high} group. PLSDA analyses clearly separated these two groups (Figure 6D), with the low-risk group largely associated with features (Figure 6E positive loadings) that mark high IgG responses against the V1V2A scaffold as well as the V1V2-169K scaffold, corresponding to Ab responses against the viral variant able to evade the vaccine response among the infected vaccinees (Rolland et al., 2012).

Correlation networks further pointed to distinct profiles between the two groups. Three subnetworks were observed in the

low-risk case:controls—an independent small network of IgA features and two larger linked clusters, including (1) all IgG3 features and (2) IgG responses tethered to Ab functional features (Figure 6F). These two clusters contained a single link between an IgG3 response and IgG response directed at the same V1V2C scaffold, which was linked to all other V1V2 scaffold responses. Again, this suggests that the IgG3 response may be a surrogate of a highly functional IgG1 response more directly involved in modulating Ab functionality. Conversely, the risk group exhibited five clusters (Figure 6G), of which four were small groups that appeared to form relationships independent of all the IgG3 features. One of the small clusters, separate from IgG3 and all functions, included several IgG V1V2 responses, highlighting a unique structure of the humoral response among the risk group. By contrast, all IgG3 features were tightly interconnected and directly tethered to IgG features and the primary ADCC and neutralization results but not to the secondary ADCC features.

These findings indicate a mis-coordinated IgG/IgG3 V1V2 response largely separate from Ab function in the vaccinees who went on to become infected, whereas IgG/IgG3 V1V2 responses were well integrated within the network profile in vaccinees with reduced correlates of risk (i.e., IgG V1V2^{high}/IgA^{low}). Even though many of the desirable features—in particular, poly-functional responses identified in Figure 5—were not available for analysis, these data highlight the IgG V1V2 responses that likely drive protective immunity.

DISCUSSION

Because the humoral immune response consists of waves of B cell responses that progressively induce higher affinity, broadly targeting, and functionally enhanced complexes of Abs poised to eliminate a pathogen, we aimed to develop a multivariate approach that could capture the complexity of interactions between Abs at unprecedented depths. The Systems Serology approach described here not only identified features reported in previous correlates analyses, including elevated IgG3 responses in RV144 (Chung et al., 2014b; Yates et al., 2014) and Ab binding to V1V2 (Zolla-Pazner et al., 2014), but also pointed to largely indirect connections between V1V2 IgG or IgG3 responses and Ab function (ADCC, ADCP, and ADCD) in vaccinees (Figure 4B) and the low-risk RV144 case:control samples (Figure 6F). Instead, vaccine-specific IgG1 responses were largely directly tethered to Ab function (Figure 4B). This suggests that the IgG3 "protective" signatures may either represent a surrogate of an effective Ab response or only contribute in combination with multiple other Ab features (e.g., IgG1) to induce antiviral activity. Along these lines, while depletion of IgG3 Abs from RV144 vaccinees resulted in a significant loss of ADCP and ADCC activity, the activity was not completely depleted with the removal of this subclass of Abs (Chung et al., 2014b), suggesting that IgG3 Abs alone do not mediate the activity in polyclonal R144 sera and that function was also mediated by Abs remaining in the depleted purified IgGs. Therefore, the induction of IgG3 responses in RV144 may mark the coordinated production of highly functional IgG1 responses that may be functionally enhanced through altered IgG1 glycosylation, known to impact Fc-receptor affinity (Chung et al., 2014a), rather than subclass

selection differences alone (Chung et al., 2014a; Hristodorov et al., 2013). Thus, together with the IgG3 Abs, these IgG1 Abs may form highly functional immune complexes that are able to rapidly and effectively clear the virus or infected cells.

Interestingly, while vaccine-induced IgA responses were associated with enhanced risk of infection (Haynes et al., 2012), IgA emerged as an antagonist of the IgG V1V2 (Figures 5A and 5B) response but not the IgG3 response in the PLSDA analyses (Figures 5C and 5D). Furthermore, IgA responses were not connected to any of the subnetworks containing functional responses identified in the network analyses, suggesting that IgA responses may serve as a marker of a deregulated or less functional humoral immune response rather than a direct antagonist of protective humoral immune responses. Thus, while it is certain that pre-incubation of Ag with IgA monoclonal Abs may prevent IgG1 and IgG3 monoclonal Ab binding (Tomaras et al., 2013), it does not appear that these responses were directly co-induced (Figure 6). Moreover, given that the infected vaccinees exhibited both the lowest and the highest levels of IgA responses (Figure 6B), it is unlikely that IgA responses directly contributed to impaired humoral immune protection. Likewise, monoclonal therapeutics generated as IgAs exhibit potent cytotoxicity and clearance of tumor targets through Fcα receptors expressed on effector cells (e.g., neutrophils and macrophages) (Black et al., 1996; Dechant and Valerius, 2001) and have been recently linked to protection from simian-HIV (SHIV) challenge (Watkins et al., 2013). Thus, future studies may aim to define the vaccine strategies that most effectively co-select a highly functional blood IgG response and a highly effective IgA response that may collectively prevent infection at the portal of entry.

Protection from infectious diseases like HIV will likely require the targeted containment of viral replication/dissemination at the site of infection. Along these lines, HIV is transmitted across mucosal barriers, where FcγR2-expressing monocytes/macrophages are abundant (Brown and Mattapallil, 2014; Zigmond and Jung, 2013). Moreover, ADCP activity was present in the RV144, VAX003, and IPCAVD001 networks (Figure 4) but was not observed in the HVTN204 network (Figure 4C) that was highly skewed to the elicitation of NK-cell-mediated activities. Conversely, ADCP was tightly tethered within the RV144 and IPCAVD001 networks (Figure 4), was enhanced in the high V1V2 IgG3/IgG1 RV144 vaccinees (Figure 5), and was previously associated with protection in NHP (Barouch et al., 2013b). Thus these results raise the possibility that ADCP may represent a critical function, within polyfunctional Ab profiles, that is required for protection from mucosal transmission.

Beyond HIV, these vaccine-profiling approaches have broad applications and can aid in vaccine design efforts against many of the deadliest global pathogens for which immune correlates of protection have yet to be elucidated. For example, recent clinical evidence suggests that Abs present in Ebola-virus-infected convalescent immune sera contribute to improved clinical outcomes in infected patients (Kreil, 2015; Lyon et al., 2014); and, recently, vesicular stomatitis virus (VSV) vaccination has been shown to drive robust humoral immune responses (Regules et al., 2015) that provide protection from infection (Henao-Restrepo et al., 2015). However, the specific mechanism(s) by which Abs provide protection remains unclear. Yet,
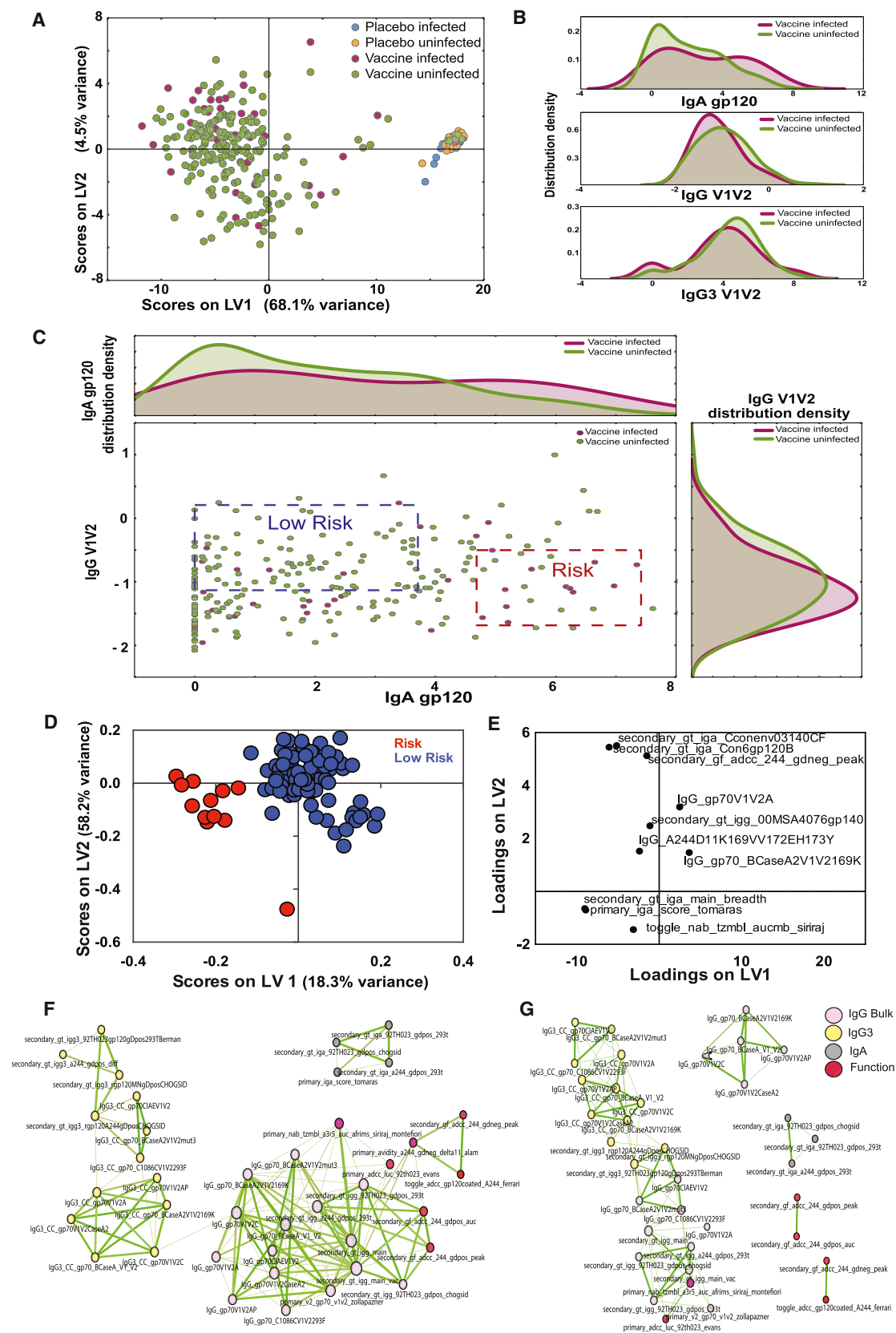
a non-neutralizing monoclonal Ab, 13c6, has been shown to provide protection from infection in an Fc-dependent manner (Olinger et al., 2012), suggesting that non-neutralizing Ab functions contribute to antiviral immunity. Thus, similar to the application of Systems Serology for the evaluation of HIV vaccine responses, the application of a Systems-Serology-guided dissection of natural humoral immune profiles that emerge in Ebola virus survivors and vaccinees may provide insights into the immunological correlates and mechanisms of protection that may help guide future vaccine efforts.

Thus, in this article, Systems Serological profiling provides a novel approach for the dissection of four HIV vaccine regimen profiles at unprecedented depths and a framework for dissecting the immune profiles that segregate with previously defined correlates of risk in efficacy studies. Systems Serology complements traditional multivariate approaches aimed at defining independent predictors of vaccine efficacy, aiding in the identification of Ab function/feature relationships that track with protective humoral immune profiles. Accordingly, these relational tools provide an additional powerful method for comparing the immune profiles of different vaccine groups/outcomes to provide greater mechanistic insights underlying the relationships of features that may contribute to immune control. While this analysis included 64 humoral features, many other features can be collected, including measures of neutralization, affinity, Fc glycosylation, and so forth. Moreover, these techniques may be expanded to examine the protective/immunopathological role of Abs in non-infectious disease settings, including malignancies and/or autoimmunity, as well as how Abs may differ among gender, ethnicity, and age. Thus, this study lays the groundwork for the evaluation, deep characterization, and comparison of polyclonal vaccine profiles for many future vaccines, for which correlates of protective immunity are still elusive.

EXPERIMENTAL PROCEDURES

Vaccine Samples

RV144 (Rerks-Ngarm et al., 2009): plasma samples from 30 vaccinated subjects at week 26 (2 weeks after the final vaccination) were provided by the Military HIV Research Program (MHRP). RV144 case:control study data were provided by the RV144 study team. Serum samples from 30 vaccinated subjects at month 30.5 (2 weeks after the final vaccination) were provided by the Global Solutions for Infectious Disease (GSID). HVTN204 (Churchyard et al., 2011): Serum samples from 30 vaccinated subjects at 2 weeks after the final vaccination were provided by the National Institute of Allergy and Infectious Diseases (NIAID) HIV Vaccine Trials Network (HVTN). IPCAVD 001 (Barouch et al., 2013a): Serum samples from 30 vaccine subjects at 2 weeks after the final

vaccination were provided by Dan Barouch. Detailed descriptions of each vaccine are included in the Supplemental Information.

Purifying Bulk IgG

IgG was purified from all vaccine plasma and serum samples using Melon Gel columns according to the manufacturer's instructions (Thermo Scientific), and the concentration was calculated using a human IgG ELISA kit (Mabtech).

Ab-Functional Profiling

The following assays were performed to functionally profile the Fc-effector functions of all vaccine Abs. In order to assess ADCP, a THP-1-based ADCP assay was performed as previously described (Ackerman et al., 2011). ADCC was assayed using a modified rapid fluorescent ADCC (RFADCC), as previously described (Gómez-Román et al., 2006); (Chung et al., 2014b). ADCD was assessed via the measurement of complement component C3b deposition on the surface of target cells. Ab-dependent NK cell degranulation and cytokine/chemokine secretion were measured using the CEM-NKr CCR5+ T-lymphoblast cell line pulsed with vaccine-specific gp120 (60 mg/ml), as previously described (Chung et al., 2014b). Detailed methods of each functional assay are described in the Supplemental Information.

Ab Biophysical Profiling

The following assays were performed to assess the biophysical profile of each of the vaccine Ab samples. Ab affinity for FCGRs was determined using surface plasmon resonance as previously described (Chung et al., 2014a), while a customized Luminex isotype assay was used to quantify the relative concentration of each Ab isotype to a panel of HIV-specific Abs. Detailed methods of each of these profiling tools are included in the Supplemental Information.

Identification of Vaccine-Specific Signatures with LASSO and PLSDA

The minimum signature of Ab features and functional parameters useful for differentiating vaccine groups were identified using the LASSO method (Tibshirani, 1997) and implemented using MATLAB software (version 2014a, MathWorks). PLSDA (Arnold et al., 2015; Lau et al., 2011) assessed the predictive ability of LASSO-selected biomarkers for classifying vaccine groups. A detailed description of validation and quality control for this analysis is included in the Supplemental Information.

Network Interactions

Networks were constructed based on the pairwise correlation coefficients between all biophysical features and functional responses. Edges between nodes are weighted using significant correlation coefficients, ρ_{ij}, after correcting for multiple comparisons (Benjamini-Hochberg q value < 0.05, testing the hypothesis of zero correlation) as follows:

$$A_{ij} = \rho_{ij}^{\alpha}$$

with $\alpha = 6$.

To assess the significance of the variable groupings observed in the network, we calculated the network clustering coefficient for the original

Figure 6. Defining Novel Signatures of Protection in the RV144 Case:Control Data
(A) The PLSDA shows the distribution of all case:control data, including all infected and uninfected placebos as well as infected and uninfected vaccinees using 101 humoral features (described in Table S3). LV1 accounted for 68.1% of all variance, separating most placebos from the vaccines, while LV2 only contributed to 4.5% of the variance.
(B) Further insights into the distribution of IgA gp120, IgG V1V2, and IgG3 V1V2 levels were analyzed using histograms demonstrating unique multi-modal differences in feature distribution among the infected and uninfected vaccinees.
(C) The scatterplot, in the central panel, represents the bivariate distribution of IgA gp120 and IgG V1V2 in the vaccines and is framed by the histogram distributions for unidimensional reference. The blue and red dash-lined boxes represent quadrants within the data that constitute the fewest cases:controls (low risk, blue) or the highest ratio of cases:controls (high risk, red).
(D and E) LASSO and PLSDA identified nine features that split low- and high-risk profile separation with 97.8% accuracy in cross-validation. Together, LV1 and LV2 captured 70.4% of the X variance and 30.1% of the Y variance, respectively.
(F and G) Correlation networks were generated for both the low-risk (F) and high-risk (G) groups.
See also Figures S2 and S3 and Table S3.

network and for 100 randomized networks. Random networks are generated by randomly swapping edges while preserving the degree of all nodes (degree-preserving edge shuffle) (Figure S2).

RV144 Case:Control Study Data Processing

RV144 case:control study data included results from 281 patients, including 101 Ab features and functional parameters. Specific features used within this analysis are documented in Table S3. Subjects were categorized into four groups including: placebo infected, placebo uninfected, vaccine infected, and vaccine uninfected for all analyses. Because IgG3 and IgG V1V2 levels were highly correlated (Figure S3), vaccinees were classified based on their IgG V1V2 and IgA levels. A high-risk or low-risk group was defined as the region of the IgG V1V2 versus IgA plot that contained the fewest cases or the fewest controls, respectively, in a mutually exclusive manner. The percentage difference between infected versus uninfected vaccinees was defined as

$$P = \left(\frac{I_r}{I_t} - \frac{U_r}{U_t} \right) \times 100,$$

where, for any given region, r, the percentage of infected people, I_r, over the infected population, I_t, was calculated, as well as for uninfected individuals. The enriched region, with the highest P was defined as the high-risk group, whereas the region with the lowest P was defined as the low-risk group. Fisher's exact test was used to estimate the significance of the enriched region from a null hypothesis.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.027.

AUTHOR CONTRIBUTIONS

Project planning was performed by A.W.C., M.P.K., K.B.A., A.E.M., D.A.L., W.H.Y., and G.A. Experimental work was performed by A.W.C., M.K.S., C.L., A.E.M., and M.H. Data analysis was performed by A.W.C., M.P.K., K.B.A., L.J.D., and W.H.Y. Vaccine trials were designed and conducted by H. Schuitemaker, M.G.P., L.R.B., J.H.K., N.L.M., and D.H.B. Manuscript composition was performed by A.W.C., M.P.K., K.B.A., W.H.Y., T.J.S., N.F., H. Streeck, M.E.A., M.J.M., H. Schuitemaker, D.A.L., and G.A.

ACKNOWLEDGMENTS

REFERENCES

Ackerman, M.E., and Alter, G. (2013). Opportunities to exploit non-neutralizing HIV-specific antibody activity. Curr. HIV Res. *11*, 365–377.

Ackerman, M.E., Moldt, B., Wyatt, R.T., Dugast, A.S., McAndrew, E., Tsoukas, S., Jost, S., Berger, C.T., Sciaranghella, G., Liu, Q., et al. (2011). A robust, high-throughput assay to determine the phagocytic activity of clinical antibody samples. J. Immunol. Methods *366*, 8–19.

Arnold, K.B., Burgener, A., Birse, K., Romas, L., Dunphy, L.J., Shahabi, K., Abou, M., Westmacott, G.R., McCorrister, S., Kwatampora, J., et al. (2015). Increased levels of inflammatory cytokines in the female reproductive tract are associated with altered expression of proteases, mucosal barrier proteins, and an influx of HIV-susceptible target cells. Mucosal Immunol. Published online June 24, 2015. http://dx.doi.org/10.1038/mi.2015.51.

Barouch, D.H., Liu, J., Peter, L., Abbink, P., Iampietro, M.J., Cheung, A., Alter, G., Chung, A., Dugast, A.S., Frahm, N., et al. (2013a). Characterization of humoral and cellular immune responses elicited by a recombinant adenovirus serotype 26 HIV-1 Env vaccine in healthy adults (IPCAVD 001). J. Infect. Dis. *207*, 248–256.

Barouch, D.H., Stephenson, K.E., Borducchi, E.N., Smith, K., Stanley, K., McNally, A.G., Liu, J., Abbink, P., Maxfield, L.F., Seaman, M.S., et al. (2013b). Protective efficacy of a global HIV-1 mosaic vaccine against heterologous SHIV challenges in rhesus monkeys. Cell *155*, 531–539.

Barouch, D.H., Alter, G., Broge, T., Linde, C., Ackerman, M.E., Brown, E.P., Borducchi, E.N., Smith, K.M., Nkolola, J.P., Liu, J., et al. (2015). HIV-1 vaccines. Protective efficacy of adenovirus/protein vaccines against SIV challenges in rhesus monkeys. Science *349*, 320–324.

Black, K.P., Cummins, J.E., Jr., and Jackson, S. (1996). Serum and secretory IgA from HIV-infected individuals mediate antibody-dependent cellular cytotoxicity. Clin. Immunol. Immunopathol. *81*, 182–190.

Bournazos, S., Klein, F., Pietzsch, J., Seaman, M.S., Nussenzweig, M.C., and Ravetch, J.V. (2014). Broadly neutralizing anti-HIV-1 antibodies require Fc effector functions for in vivo activity. Cell *158*, 1243–1253.

Brown, D., and Mattapallil, J.J. (2014). Gastrointestinal tract and the mucosal macrophage reservoir in HIV infection. Clin. Vaccine Immunol. *21*, 1469–1473.

Chaudhuri, J., and Alt, F.W. (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat. Rev. Immunol. *4*, 541–552.

Chaudhuri, J., Basu, U., Zarrin, A., Yan, C., Franco, S., Perlot, T., Vuong, B., Wang, J., Phan, R.T., Datta, A., et al. (2007). Evolution of the immunoglobulin heavy chain class switch recombination mechanism. Adv. Immunol. *94*, 157–214.

Chung, A.W., and Alter, G. (2014). Dissecting the antibody constant region protective immune parameters in HIV infection. Future Virol. *9*, 397–414.

Chung, A.W., Crispin, M., Pritchard, L., Robinson, H., Gorny, M.K., Yu, X., Bailey-Kellogg, C., Ackerman, M.E., Scanlan, C., Zolla-Pazner, S., and Alter, G. (2014a). Identification of antibody glycosylation structures that predict monoclonal antibody Fc-effector function. AIDS *28*, 2523–2530.

Chung, A.W., Ghebremichael, M., Robinson, H., Brown, E., Choi, I., Lane, S., Dugast, A.S., Schoen, M.K., Rolland, M., Suscovich, T.J., et al. (2014b). Polyfunctional Fc-effector profiles mediated by IgG subclass selection distinguish RV144 and VAX003 vaccines. Sci. Transl. Med. *6*, 228ra38.

Churchyard, G.J., Morgan, C., Adams, E., Hural, J., Graham, B.S., Moodie, Z., Grove, D., Gray, G., Bekker, L.-G., McElrath, M.J., et al.; NIAID HIV Vaccine Trials Network (2011). A phase IIA randomized clinical trial of a multiclade HIV-1 DNA prime followed by a multiclade rAd5 HIV-1 vaccine boost in healthy adults (HVTN204). PLoS ONE *6*, e21225.

Dechant, M., and Valerius, T. (2001). IgA antibodies for cancer therapy. Crit. Rev. Oncol. Hematol. *39*, 69–77.

DiLillo, D.J., Tan, G.S., Palese, P., and Ravetch, J.V. (2014). Broadly neutralizing hemagglutinin stalk-specific antibodies require FcγR interactions for protection against influenza virus in vivo. Nat. Med. 20, 143–151.

Gómez-Román, V.R., Florese, R.H., Patterson, L.J., Peng, B., Venzon, D., Aldrich, K., and Robert-Guroff, M. (2006). A simplified method for the rapid fluorometric assessment of antibody-dependent cell-mediated cytotoxicity. J. Immunol. Methods 308, 53–67.

Haynes, B.F., Gilbert, P.B., McElrath, M.J., Zolla-Pazner, S., Tomaras, G.D., Alam, S.M., Evans, D.T., Montefiori, D.C., Karnasuta, C., Sutthent, R., et al. (2012). Immune-correlates analysis of an HIV-1 vaccine efficacy trial. N. Engl. J. Med. 366, 1275–1286.

Henao-Restrepo, A.M., Longini, I.M., Egger, M., Dean, N.E., Edmunds, W.J., Camacho, A., Carroll, M.W., Doumbia, M., Draguez, B., Duraffour, S., et al. (2015). Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: interim results from the Guinea ring vaccination cluster-randomised trial. Lancet 386, 857–866.

Hessell, A.J., Hangartner, L., Hunter, M., Havenith, C.E., Beurskens, F.J., Bakker, J.M., Lanigan, C.M., Landucci, G., Forthal, D.N., Parren, P.W., et al. (2007). Fc receptor but not complement binding is important in antibody protection against HIV. Nature 449, 101–104.

Hristodorov, D., Fischer, R., and Linden, L. (2013). With or without sugar? (A) glycosylation of therapeutic antibodies. Mol. Biotechnol. 54, 1056–1068.

Jegerlehner, A., Schmitz, N., Storni, T., and Bachmann, M.F. (2004). Influenza A vaccine based on the extracellular domain of M2: weak protection mediated via antibody-dependent NK cell activity. J. Immunol. 172, 5598–5605.

Joos, C., Marrama, L., Polson, H.E., Corre, S., Diatta, A.M., Diouf, B., Trape, J.F., Tall, A., Longacre, S., and Perraut, R. (2010). Clinical protection from falciparum malaria correlates with neutrophil respiratory bursts induced by merozoites opsonized with human serum antibodies. PLoS ONE 5, e9871.

Kohl, S., and Loo, L.S. (1982). Protection of neonatal mice against herpes simplex virus infection: probable in vivo antibody-dependent cellular cytotoxicity. J. Immunol. 129, 370–376.

Kohl, S., Loo, L.S., and Pickering, L.K. (1981). Protection of neonatal mice against herpes simplex viral infection by human antibody and leukocytes from adult, but not neonatal humans. J. Immunol. 127, 1273–1275.

Kreil, T.R. (2015). Treatment of Ebola virus infection with antibodies from reconvalescent donors. Emerg. Infect. Dis. 21, 521–523.

Lau, K.S., Juchheim, A.M., Cavaliere, K.R., Philips, S.R., Lauffenburger, D.A., and Haigis, K.M. (2011). In vivo systems analysis identifies spatial and temporal aspects of the modulation of TNF-α-induced apoptosis and proliferation by MAPKs. Sci. Signal. 4, ra16.

Lyon, G.M., Mehta, A.K., Varkey, J.B., Brantly, K., Plyler, L., McElroy, A.K., Kraft, C.S., Towner, J.S., Spiropoulou, C., Ströher, U., et al.; Emory Serious Communicable Diseases Unit (2014). Clinical care of two patients with Ebola virus disease in the United States. N. Engl. J. Med. 371, 2402–2409.

Olinger, G.G., Jr., Pettitt, J., Kim, D., Working, C., Bohorov, O., Bratcher, B., Hiatt, E., Hume, S.D., Johnson, A.K., Morton, J., et al. (2012). Delayed treatment of Ebola virus infection with plant-derived monoclonal antibodies provides protection in rhesus macaques. Proc. Natl. Acad. Sci. USA 109, 18030–18035.

Osier, F.H., Feng, G., Boyle, M.J., Langer, C., Zhou, J., Richards, J.S., McCallum, F.J., Reiling, L., Jaworowski, A., Anders, R.F., et al. (2014). Opsonic phagocytosis of Plasmodium falciparum merozoites: mechanism in human immunity and a correlate of protection against malaria. BMC Med. 12, 108.

Pitisuttithum, P., Gilbert, P., Gurwith, M., Heyward, W., Martin, M., van Griensven, F., Hu, D., Tappero, J.W., and Choopanya, K.; Bangkok Vaccine Evaluation Group (2006). Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. J. Infect. Dis. 194, 1661–1671.

Pulendran, B., and Ahmed, R. (2011). Immunological mechanisms of vaccination. Nat. Immunol. 12, 509–517.

Regules, J.A., Beigel, J.H., Paolino, K.M., Voell, J., Castellano, A.R., Muñoz, P., Moon, J.E., Ruck, R.C., Bennett, J.W., Twomey, P.S., et al.; rVSVΔG-ZEBOV-GP Study Group (2015). A recombinant vesicular stomatitis virus Ebola vaccine—preliminary report. N. Engl. J. Med. Published online April 1, 2015. http://dx.doi.org/10.1056/NEJMoa1414216.

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., de Souza, M., Adams, E., et al.; MOPH-TAVEG Investigators (2009). Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N. Engl. J. Med. 361, 2209–2220.

Rolland, M., Edlefsen, P.T., Larsen, B.B., Tovanabutra, S., Sanders-Buell, E., Hertz, T., deCamp, A.C., Carrico, C., Menis, S., Magaret, C.A., et al. (2012). Increased HIV-1 vaccine efficacy against viruses with genetic signatures in Env V2. Nature 490, 417–420.

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. Stat. Med. 16, 385–395.

Tomaras, G.D., Ferrari, G., Shen, X., Alam, S.M., Liao, H.X., Pollara, J., Bonsignori, M., Moody, M.A., Fong, Y., Chen, X., et al. (2013). Vaccine-induced plasma IgA specific for the C1 region of the HIV-1 envelope blocks binding and effector function of IgG. Proc. Natl. Acad. Sci. USA 110, 9019–9024.

Warfield, K.L., Swenson, D.L., Olinger, G.G., Kalina, W.V., Aman, M.J., and Bavari, S. (2007). Ebola virus-like particle-based vaccine protects nonhuman primates against lethal Ebola virus challenge. J. Infect. Dis. 196 (Suppl 2), S430–S437.

Watkins, J.D., Sholukh, A.M., Mukhtar, M.M., Siddappa, N.B., Lakhashe, S.K., Kim, M., Reinherz, E.L., Gupta, S., Forthal, D.N., Sattentau, Q.J., et al.; CAVD Project Group (2013). Anti-HIV IgA isotypes: differential virion capture and inhibition of transcytosis are linked to prevention of mucosal R5 SHIV transmission. AIDS 27, F13–F20.

White, A.L., Beers, S.A., and Cragg, M.S. (2014). FcγRIIB as a key determinant of agonistic antibody efficacy. Curr. Top. Microbiol. Immunol. 382, 355–372.

Yates, N.L., Liao, H.X., Fong, Y., deCamp, A., Vandergrift, N.A., Williams, W.T., Alam, S.M., Ferrari, G., Yang, Z.Y., Seaton, K.E., et al. (2014). Vaccine-induced Env V1-V2 IgG3 correlates with lower HIV-1 infection risk and declines soon after vaccination. Sci. Transl. Med. 6, 228ra39.

Zigmond, E., and Jung, S. (2013). Intestinal macrophages: well educated exceptions from the rule. Trends Immunol. 34, 162–168.

Zolla-Pazner, S., deCamp, A., Gilbert, P.B., Williams, C., Yates, N.L., Williams, W.T., Howington, R., Fong, Y., Morris, D.E., Soderberg, K.A., et al. (2014). Vaccine-induced IgG antibodies to V1V2 regions of multiple HIV-1 subtypes correlate with decreased risk of HIV-1 infection. PLoS ONE 9, e87572.

Cell

5′ UTR m6A Promotes Cap-Independent Translation

Graphical Abstract
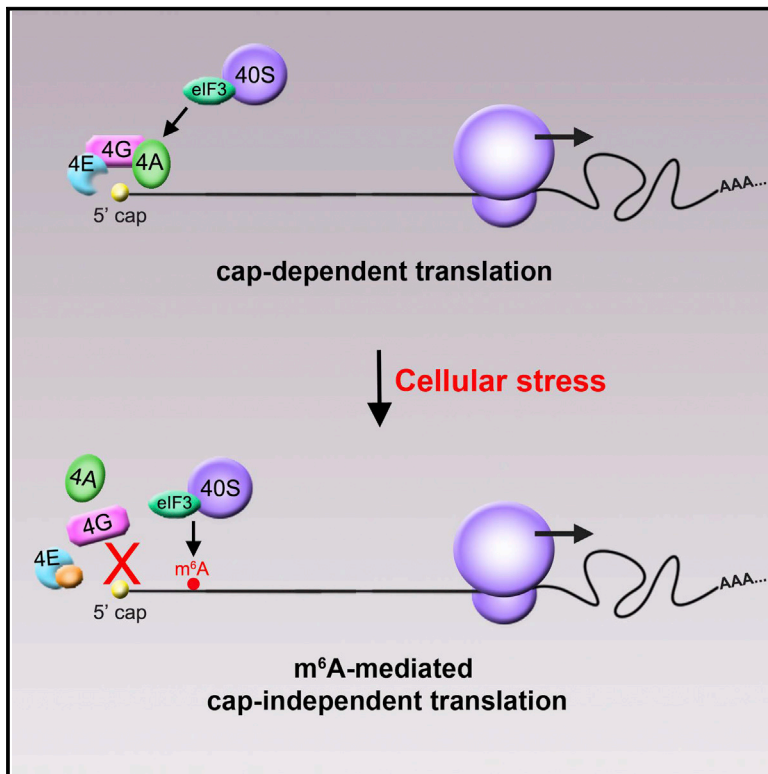
Authors

Kate D. Meyer, Deepak P. Patil, Jun Zhou, ..., Tatyana V. Pestova, Shu-Bing Qian, Samie R. Jaffrey

Correspondence

srj2003@med.cornell.edu

In Brief

N^6-methyladenosine (m6A) residues within the 5′ UTR of mRNAs promote translation initiation through a mechanism that does not require the 5′ cap or cap-binding proteins. Diverse cellular stresses selectively increase the levels of m6A within 5′ UTRs, suggesting that 5′ UTR m6A is important for mediating stress-induced translational responses.

Highlights

- m6A residues within the 5′ UTR promote cap-independent translation

- Translation of cellular mRNAs is increased by the presence of m6A within the 5′ UTR

- Heat shock induces *Hsp70* translation in an m6A-dependent manner

- Diverse cellular stresses increase 5′ UTR adenosine methylation

Accession Numbers

GSE73405

CellPress

Article

5′ UTR m^{6}A Promotes Cap-Independent Translation

Kate D. Meyer,[1] Deepak P. Patil,[1] Jun Zhou,[2] Alexandra Zinoviev,[3] Maxim A. Skabkin,[3] Olivier Elemento,[4,5] Tatyana V. Pestova,[3] Shu-Bing Qian,[2] and Samie R. Jaffrey[1,*]

[1]Department of Pharmacology, Weill Medical College, Cornell University, New York, NY 10065, USA
[2]Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853, USA
[3]Department of Cell Biology, SUNY Downstate Medical Center, 450 Clarkson Avenue, Brooklyn, NY 11203, USA
[4]Department of Physiology and Biophysics, Weill Medical College, Cornell University, New York, NY 10065, USA
[5]HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Medical College, Cornell University, New York, NY 10065, USA
*Correspondence: srj2003@med.cornell.edu
http://dx.doi.org/10.1016/j.cell.2015.10.012

SUMMARY

Protein translation typically begins with the recruitment of the 43S ribosomal complex to the 5′ cap of mRNAs by a cap-binding complex. However, some transcripts are translated in a cap-independent manner through poorly understood mechanisms. Here, we show that mRNAs containing N^6-methyladenosine (m^{6}A) in their 5′ UTR can be translated in a cap-independent manner. A single 5′ UTR m^{6}A directly binds eukaryotic initiation factor 3 (eIF3), which is sufficient to recruit the 43S complex to initiate translation in the absence of the cap-binding factor eIF4E. Inhibition of adenosine methylation selectively reduces translation of mRNAs containing 5′UTR m^{6}A. Additionally, increased m^{6}A levels in the Hsp70 mRNA regulate its cap-independent translation following heat shock. Notably, we find that diverse cellular stresses induce a transcriptome-wide redistribution of m^{6}A, resulting in increased numbers of mRNAs with 5′ UTR m^{6}A. These data show that 5′ UTR m^{6}A bypasses 5′ cap-binding proteins to promote translation under stresses.

INTRODUCTION

For most cellular mRNAs, the first step of mRNA translation involves recognition of the 5′ 7-methylguanosine (m^{7}G) cap by eukaryotic initiation factor 4E (eIF4E), which is a subunit of the heterotrimeric eIF4F complex. 5′ cap-bound eIF4F then recruits the small (40S) ribosomal subunit associated with various translation initiation factors, enabling efficient translation of eukaryotic mRNAs.

However, some mRNAs are translated in a cap-independent manner. These capped mRNAs do not require eIF4E and are translated under basal cellular conditions, as well as conditions in which eIF4E activity is compromised, such as cellular stress states, viral infection, and diseases such as cancer (Stoneley and Willis, 2004). Although viral mRNAs can exhibit cap-independent translation due to the presence of highly structured internal ribosome entry site (IRES) motifs in the 5′ UTR, corre-

spondingly complex structures are rarely found in eukaryotic mRNAs undergoing cap-independent translation (Stoneley and Willis, 2004). Thus, the mechanism of cap-independent translation in cellular mRNAs remains poorly understood.

A feature of many eukaryotic mRNAs is N^6-methyladenosine (m^{6}A), a reversible base modification seen in the 3′ UTR, coding sequence, and 5′ UTR (Dominissini et al., 2012; Meyer et al., 2012). Although the function of m^{6}A in 3′UTRs has been explored (Wang et al., 2014a, 2014b, 2015), the function of m^{6}A in 5′ UTRs remains unknown. Here, we show that m^{6}A in the 5′ UTR functions as an alternative to the 5′ cap to stimulate mRNA translation. Using both in vitro reconstitution approaches and translation assays in cellular lysates deficient in eIF4E activity, we define a unique translation initiation mechanism that does not require the 5′ cap. We show that the m^{6}A in the 5′ UTR can bind eukaryotic initiation factor 3 (eIF3). Transcriptome-wide ribosome profiling analysis indicates that the translation of 5′ UTR m^{6}A-containing mRNAs is reduced upon depletion of the m^{6}A methyltransferase, METTL3, while mRNAs containing m^{6}A elsewhere within the transcript fail to show this effect. The importance of 5′ UTR m^{6}A residues for cellular mRNA translation is demonstrated by both ribosome profiling analysis and detection of changes to global m^{6}A distribution in 5′UTRs in response to cellular stress. Thus, 5′ UTR m^{6}A residues are linked to cellular stress states and provide a mechanism to bypass the m^{7}G cap requirement for mRNA translation, enabling a cap-independent mode of translation initiation.

RESULTS

Ribosomal Initiation Complexes Assemble on m^{6}A-Containing mRNAs Independently of the Cap-Binding Protein eIF4E

Although m^{6}A is predominantly localized near stop codons and in 3′ UTRs in several thousand mRNAs, hundreds of cellular mRNAs contain m^{6}A within their 5′ UTR (Linder et al., 2015; Meyer et al., 2012), and the function of these m^{6}A residues is unknown. Since the 5′ UTR is important in regulating translation initiation, we considered the possibility that 5′ UTR-localized m^{6}As might influence this process. On most eukaryotic mRNAs, translation begins with assembly of a 43S preinitiation complex, comprising a 40S ribosomal subunit, a eukaryotic initiation factor 2 (eIF2)-GTP/Met-tRNA$_i^{Met}$ ternary complex, and eIFs 3, 1,

Cell

and 1A (Jackson et al., 2010). 43S complexes are typically recruited to mRNA by a cap-binding complex, eIF4F. eIF4F consists of three subunits: eIF4E, which binds the m^7G 5′ cap; eIF4A, an RNA helicase; and eIF4G, a scaffold that also binds eIF3, thereby recruiting the 43S complex. After attachment, 43S complexes scan to the initiation codon, where they form 48S initiation complexes (Jackson et al., 2010).

To investigate the effect of m^6A on translation initiation, we used toeprinting, an approach for reconstituting assembly of 48S complexes on mRNA. In toeprinting, ribosomal complexes are assembled on mRNA 5′ UTRs using purified translational components (40S subunits, initiation factors and Met-tRNA$_i^{Met}$) (Pestova and Kolupaeva, 2002). Formation of the 48S complex at the start codon is then monitored by reverse transcriptase-mediated extension of a [^{32}P]-labeled primer annealed to ribosome-bound mRNA. cDNA synthesis is arrested by the 40S ribosome subunit, yielding characteristic toeprints at its leading edge, +15–17 nt downstream of the initiation codon. This assay can identify the initiation factors and sequence features of 5′ UTRs that are required for initiation and has been used in mechanistic studies of viral IRESs (Pestova and Hellen, 2003).

To test the role of m^6A in 48S complex formation, we performed toeprinting with 5′-capped mRNAs comprising the 54-nt-long β-globin 5′ UTR followed by a short coding sequence, stop codon, and 3′ UTR. Consistent with previous studies (Pestova and Kolupaeva, 2002), 48S initiation complexes were detected at the start codon of A-containing mRNA in the presence of the complete set of eIFs (1, 1A, 2, 3, 4A, 4B, and 4F), and omission of group 4 eIFs nearly abrogated 48S complex formation (Figure 1A, compare lanes 2 and 4). This is consistent with the known role for the eIF4 cap-binding complex in recruiting the 43S complex to mRNA (Gingras et al., 1999).

When we used mRNAs that were in vitro transcribed to contain m^6A, we found that 48S complexes readily assembled after addition of the complete set of eIFs, as was seen with unmethylated mRNA. However, unlike the unmethylated mRNA, 48S complexes formed on m^6A-containing mRNA even in the absence of group 4 eIFs (Figure 1A). Thus, initiation on m^6A-containing mRNA is distinct from initiation on mRNA lacking m^6A and does not require the eIF4 cap-binding complex.

To further establish the factor requirements for initiation on m^6A-containing mRNA, we selectively omitted each initiation factor and performed toeprinting. These experiments show that efficient initiation on m^6A-containing mRNA only requires the presence of eIFs1, 1A, 2, 3, and the 40S subunit (Figures 1B and 1C). 48S complexes that formed on m^6A-containing mRNA in the absence of group 4 eIFs were functional, as addition of the 60S ribosomal subunit, Σaa-tRNAs, and factors required for subunit joining and elongation (eIF5, eIF5B, eEF2, and eEF1H) resulted in formation of 80S ribosomes that underwent efficient elongation and yielded pre-termination complexes at the stop codon (Figure 1B). Thus, translation-competent 48S complexes can form on m^6A-containing mRNA in the absence of eIF4E.

m6A Enables Translation in a 5′ Cap-Independent Manner in Cell-Free Extracts

We next asked if m^6A induces eIF4E-independent translation in cell-free extracts. To investigate this, we used a HeLa extract that has low eIF4E activity (Mikami et al., 2006) (Figures S1A and S1B) and thus provides an ideal system for studying eIF4E-independent translation. Indeed, addition of a capped, nonmethylated luciferase-encoding mRNA containing the β-globin 5′ UTR to the HeLa extract did not produce measureable luciferase activity unless eIF4E was added (Figure 2A). Thus, cap-dependent translation in this extract is dependent on exogenous eIF4E.

We next used HeLa extracts to determine if transcripts containing m^6A require eIF4E. In contrast to the mRNA containing exclusively A, 5′-capped mRNA containing 50% m^6A was readily translated even in the absence of added eIF4E (Figure 2A). Furthermore, addition of 1 mM m^7GpppG, a cap analog that sequesters cap-binding proteins (Ray et al., 2006), abolished translation of 5′-capped, A-containing mRNA but had no effect on m^6A-containing mRNA (Figure 2B). Lastly, A-containing mRNA synthesized without a cap was not translated, whereas m^6A-containing, uncapped mRNA was readily translated (Figure 2C). The increased translation of m^6A-containing mRNA in these experiments was not due to increased stability of m^6A-containing mRNA, as RT-qPCR and radiolabeled mRNA stability measurements indicated similar levels of A- and m^6A-containing luciferase mRNA after incubation with HeLa extracts (Figures S1C–S1F). Collectively, these data indicate that translation of m^6A-containing mRNA exhibits marked independence of the 5′ cap and eIF4E.

A Single m6A Is Sufficient to Induce Cap-Independent Translation

Since the mRNAs used in the in vitro translation assays have m^6A throughout the transcript, it is unclear if the translational effects are due to m^6A in the 5′ UTR or elsewhere in the mRNA. To determine the contributions of specific m^6A residues to cap-independent translation, we examined mRNAs that only contain m^6A in the coding sequence. Uncapped, luciferase-encoding mRNAs that contained zero m^6A residues within the 5′ UTR showed no translation, indicating that m^6A residues in the coding sequence are unable to induce cap-independent translation (Figure 2D). However, addition of a single m^6A residue at the beginning, middle, or end of the 5′ UTR was sufficient to markedly induce cap-independent translation (Figure 2D).

To determine if a single 5′ UTR m^6A residue can promote cap-independent translation, we used uncapped luciferase-encoding mRNAs that contain m^6A as the first transcribed nucleotide (Supplemental Experimental Procedures). This mRNA contains a single m^6A residue in the 5′ UTR, and the remainder of the As within the transcript are unmethylated. For mRNAs lacking m^6A, negligible luciferase synthesis was detected (Figure 2E). However, transcripts containing a single 5′ m^6A were readily translated (Figure 2E). Notably, the level of translation induced by a 5′ m^6A is less than the translation induced by a single m^6A residue located internally within the 5′ UTR, which likely reflects inefficient incorporation of m^6A at the first position of 5′ m^6A-containing transcripts (see Experimental Procedures). Collectively, these experiments indicate that a single m^6A can induce cap-independent translation.

To determine whether m^6A-mediated cap-independent translation is a specific effect caused by the presence of m^6A, we synthesized uncapped luciferase transcripts containing A,
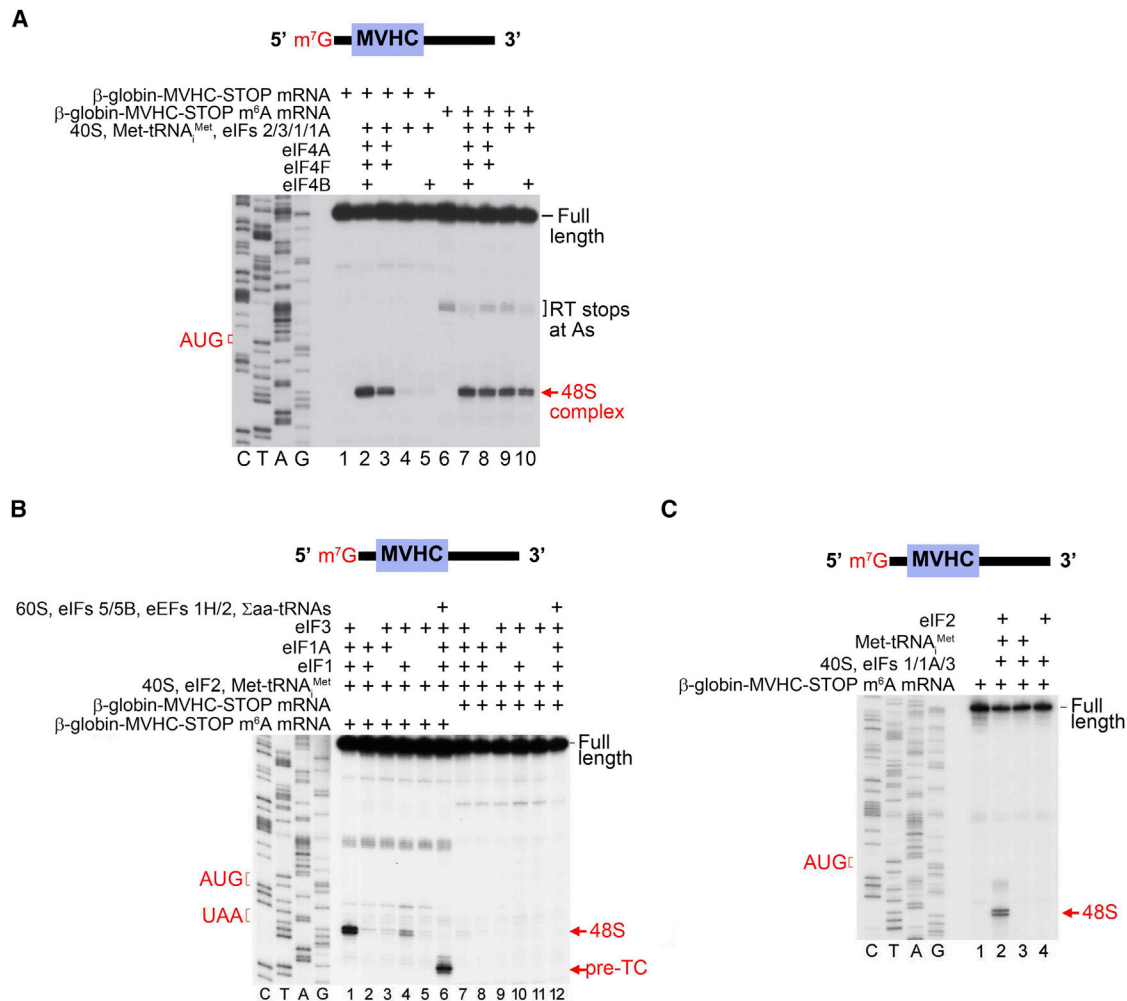
Cell

Figure 1. 5′UTR m⁶A Enables Ribosome Binding to mRNA in the Absence of Cap-Binding Proteins

(A) 5′ UTR methylation permits 48S initiation complex formation in the absence of the group 4 eIFs. In vitro transcribed, capped mRNAs encoding a MVHC tetrapeptide and containing either A or m⁶A were incubated with purified mammalian translation initiation components. Subsequent toeprinting analysis using a radiolabeled primer then revealed whether 48S initiation complexes were formed. Positions of the initiation codon, full-length cDNA, and the 48S complex are shown on the sides of the panel. Lanes C/T/A/G depict the corresponding DNA sequence. When unmethylated mRNA is used (lanes 1–5), 48S complexes are only formed when the cap-binding complex eIF4F is present (lanes 2 and 3). When eIF4F is absent, 48S complex formation on unmethylated mRNA is impaired (lanes 4 and 5). However, when mRNA with m⁶A in the 5′ UTR is used, 48S complex formation is observed even in the absence of eIF4F (lanes 9 and 10; compare to lanes 7 and 8 where eIF4F is present).

(B) eIFs1, 1A, and 3 are required for efficient m⁶A-induced cap-independent 48S complex formation. Toeprinting assays were performed as in (A) using A- or m⁶A-containing mRNAs and in the presence of various translation initiation components as indicated. m⁶A-containing mRNA exhibits robust 48S complex assembly in the absence of eIF4F, whereas A-containing mRNA does not (compare lanes 1 and 7). Efficient m⁶A-mediated 48S complex assembly is also dependent on the presence of eIFs1 and 1A, which is consistent with the known roles of these proteins in promoting scanning and AUG recognition (compare lanes 1 with lanes 2, 4, and 5). Removal of eIF3 also abolishes 48S complex assembly on m⁶A-containing mRNA (compare lanes 1 and 2), indicating that eIF3 is required for m⁶A-mediated 48S complex formation. Addition of 60S subunits, eIF5, eIF5B, eEF1H, eEF2, and aa-tRNAs resulted in the appearance of toeprints corresponding to pre-termination complexes at the stop codon, indicating that m⁶A-recruited 48S complexes are fully functional (lane 6).

(C) Omission of eIF2 from toeprinting assays results in the absence of 48S complexes (compare lanes 2 and 3), indicating that eIF2 is required for 48S complex assembly on m⁶A-containing mRNA.

m⁶A, or other modified nucleotides, such as N^1-methyladenosine, 2′-O-methyladenosine, pseudouridine, and 5-methylcytosine. In each case, there was negligible luciferase synthesis unless m⁶A was present (Figure S1G).

We next asked if the effect of m⁶A reflects impaired base pairing caused by modification of the N^6 position (Roost et al., 2015).

However, mRNA containing N^6-propargyladenosine, which contains a slightly larger modification compared to a methyl group at the N^6 position, failed to undergo cap-independent translation (Figure S1G). Thus, m⁶A-induced structural changes are unlikely to account for the cap independence conferred by m⁶A.
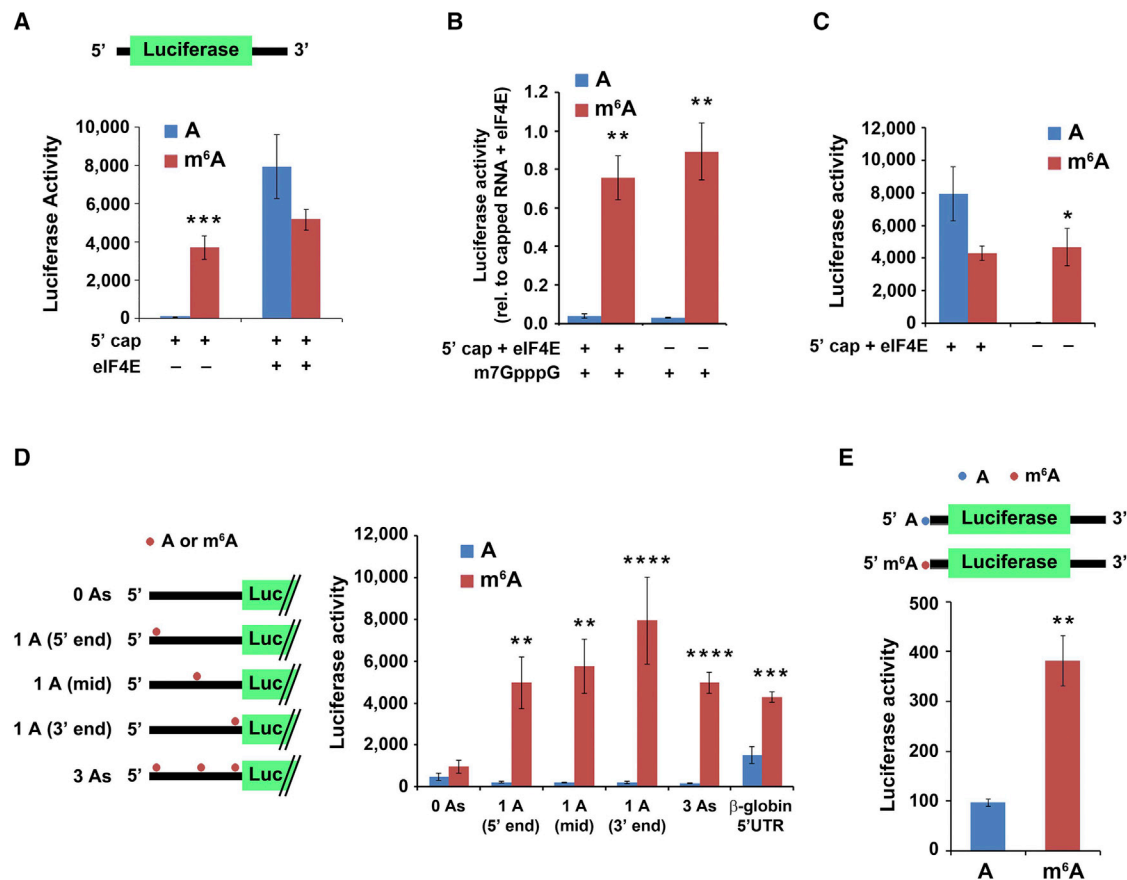
Figure 2. m6A within the 5′ UTR Enables Cap-Independent Translation of mRNA

(A) 5′ UTR m6A permits mRNA translation without the need for the cap-binding protein eIF4E. In vitro translation was performed using a HeLa cell extract mixed with luciferase-encoding, capped mRNA containing either A or m6A. Protein production was measured by quantifying luciferase activity. Cap-dependent translation is observed from both methylated and unmethylated mRNAs in the presence of eIF4E. However, when eIF4E is absent, only the m6A-containing mRNA is translated (n = 4; mean ± SD; ***p < 0.0001).

(B) Presence of a 5′ cap analog is unable to abolish m6A-induced mRNA translation. Luciferase mRNAs were translated as in (A). 1 mM free cap analog (m7GpppG) was added to sequester cap-binding proteins. Addition of m7GpppG abolishes cap-dependent translation of unmethylated mRNA (left) but is unable to abolish the cap-independent translation induced by m6A (right). Levels of luciferase activity are shown relative to capped mRNA + 10 pmole eIF4E (n = 3; mean ± SD; *p < 0.01, **p < 0.001).

(C) In vitro translation was performed using luciferase-encoding mRNA containing A or 50% m6A and with or without a 5′cap as indicated. While unmethylated, capped mRNA + 10 pmole eIF4E is robustly translated, the unmethylated, uncapped mRNA fails to be translated. However, m6A-containing mRNA is efficiently translated even when no 5′ cap is present (n = 3; mean ± SD; *p < 0.01).

(D) m6A residues in the coding sequence do not induce cap-independent translation. Uncapped, luciferase-encoding mRNAs containing either the natural β-globin 5′ UTR or a modified β-globin 5′ UTR containing either zero, one, or three A residues as indicated were used for in vitro translation assays. Translation of m6A-containing mRNA with zero A residues in the 5′ UTR was markedly diminished, indicating that coding sequence m6A residues are unable to induce cap-independent translation. However, when a single m6A was added to the 5′ UTR, the transcripts were robustly translated. Methylated 5′ UTRs with a single A near the 5′ end, the middle (mid), or near the 3′ end all showed similar levels of translation (n = 3; mean ± SD; **p < 0.001, ***p < 0.0001, ****p < 0.00001). Schematic shows the distribution of A residues within each β-globin 5′ UTR variant (the unmodified β-globin 5′ UTR contains 17 A residues).

(E) mRNA with a single m6A within the 5′ UTR and no m6As in the remainder of the transcript induces cap-independent translation. Uncapped, luciferase-encoding mRNAs, which contained either a single adenosine 5′-monophosphate (AMP) or N6-methyladenosine 5′-monophosphate (m6AMP) at the 5′ end, were used for in vitro translation. Only the m6A-containing mRNA was translated, demonstrating that a single 5′ end m6A residue is capable of inducing cap-independent translation (n = 3; mean ± SD; **p < 0.001). The reduced translation efficiency of this mRNA compared to mRNAs with internally methylated 5′ UTRs is likely due to inefficient incorporation of m6A residues at the 5′ end by T7 RNA polymerase.

See also Figure S1.

m6A-Induced Translation Initiation Occurs through a 5′ End-Dependent Mechanism

Our results indicate that m6A residues within the 5′ UTR are capable of promoting cap-independent translation. However, the majority of m6A residues are found in the coding sequence and 3′ UTR (Meyer et al., 2012). We therefore asked if these internal m6A residues can induce internal ribosome entry. To test this, we synthesized an m6A-containing β-globin mRNA in which the wild-type AUG initiation codon was removed and two new AUG triplets were introduced upstream and downstream of the native
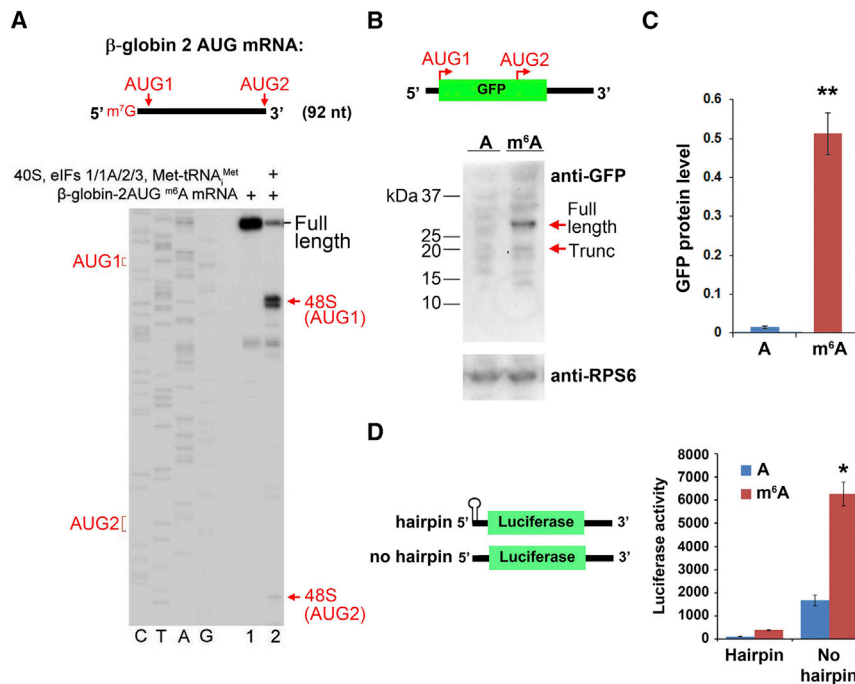
Cell

Figure 3. m⁶A-Mediated Translation Occurs through a 5′ End-Dependent Mechanism

(A) Toeprinting assays were performed using a capped, m⁶A-containing mRNA containing the β-globin 5′ UTR sequence, which was modified to include two AUG initiation codons ("AUG1" and "AUG2" in the schematic). The majority of 48S complexes were assembled at AUG1, with negligible levels of 48S complexes detected at AUG2.

(B) Uncapped, A-, or m⁶A-containing mRNAs encoding GFP were used for in vitro translation. The mRNA contains two near-kozak start codons: AUG 1 encodes the full-length GFP protein, and internally localized AUG2 encodes an in-frame truncated (∼17 kDa) protein comprising the C-terminal portion of GFP. Full-length and truncated GFP protein levels (sizes indicated by arrows) were measured by western blot. m⁶A primarily promotes translation of the full-length protein and fails to induce internal entry-mediated translation from AUG2. Levels of the ribosomal protein RPS6 are shown as a loading control.

(C) Quantification of full-length GFP protein levels in (B) shows increased protein expression of methylated mRNA versus unmethylated mRNA (n = 3; mean ± SD; **p < 0.001).

(D) The presence of a stable hairpin at the beginning of the 5′ UTR to block 5′ end entry severely attenuates m⁶A-mediated translation (n = 3; mean ± SD; *p < 0.01).

See also Figure S1.

position (Figure 3A). When this mRNA was incubated with 40S, eIFs1/1A/2/3, and Met-tRNAᵢᴹᵉᵗ, 48S complexes occurred almost exclusively at the first AUG, with very low levels of detectable 48S complex formation at the downstream AUG (Figure 3A). These data suggest that m⁶A preferentially induces translation at the first suitable start codon in the mRNA as opposed to promoting translation through an internal entry-based mechanism.

Next, we used HeLa cell lysates to in vitro translate a GFP reporter mRNA containing an internal near-Kozak AUG in addition to the natural AUG encoding full-length GFP. However, we failed to observe m⁶A-mediated translation of the ∼17 kDa product produced from the internal AUG and instead observed robust translation of the full-length protein produced from the first AUG (Figures 3B, 3C, and S1H). These results are consistent with the toeprinting experiments and suggest that m⁶A preferentially induces translation at the first acceptable start codon.

The selective use of the first AUG for translation initiation suggests a model of m⁶A-mediated initiation that involves a 5′ end-dependent scanning mechanism as opposed to internal ribosomal entry. A similar mode of initiation, which is also cap independent but shows 5′-end dependence, was recently described for mRNA containing in its 5′ UTR an eIF4G-binding viral IRES-domain (Terenin et al., 2013). Additionally, cap-independent, 5′ end-dependent mechanisms of translation initiation have previously been observed in assays using rabbit reticulocyte lysates (De Gregorio et al., 1998). To test directly whether m⁶A promotes entry through the 5′ end, we used an uncapped, luciferase-encoding mRNA that contains a stable hairpin at the extreme 5′ end of the mRNA to block 5′ end-dependent ribosome entry. We found that the presence of this hairpin markedly

reduced the robust translation of m⁶A-containing mRNA that is normally observed (Figure 3D). Thus, m⁶A-mediated initiation requires an accessible 5′-terminal end on the mRNA. Taken together, these data indicate that 5′ UTR m⁶As are distinct from classical viral IRES elements since m⁶A promotes recruitment of ribosomal preinitiation complexes to the 5′ end of mRNA, rather than enabling internal ribosome entry.

eIF3 Selectively Binds m⁶A-Containing RNA

We next asked how m⁶A is recognized to induce translation of mRNAs. The in vitro 48S reconstitution assays showed that recruitment of the 43S preinitiation complex to m⁶A-containing mRNA only requires eIFs 1, 1A, 2, and 3 and the 40S subunit. Thus, one of these components binds m⁶A.

To test which of these factors interacts with m⁶A, we used an m⁶A crosslinking assay in which a [³²P]-labeled RNA probe containing a single A or m⁶A in its naturally occurring GAC context was UV-crosslinked to each translational component. Crosslinked proteins were then detected by SDS-PAGE and autoradiography.

eIFs 1, 1A, and 2 and the 40S subunit showed equal levels of crosslinking to the A- and m⁶A-containing probes (Figures 4A and S2A). However, crosslinking of eIF3 to the m⁶A-containing probe was substantially increased compared to the A-containing probe, suggesting that this factor constitutes the major m⁶A-binding activity of the 43S complex (Figures 4A, S2B, and S2C).

The preferential binding of eIF3 to m⁶A was not affected by changing the position of the m⁶A together with its context nucleotides within the probe (Figure S2D). However, when the natural nucleotide context of m⁶A was changed from GAC to UAC or
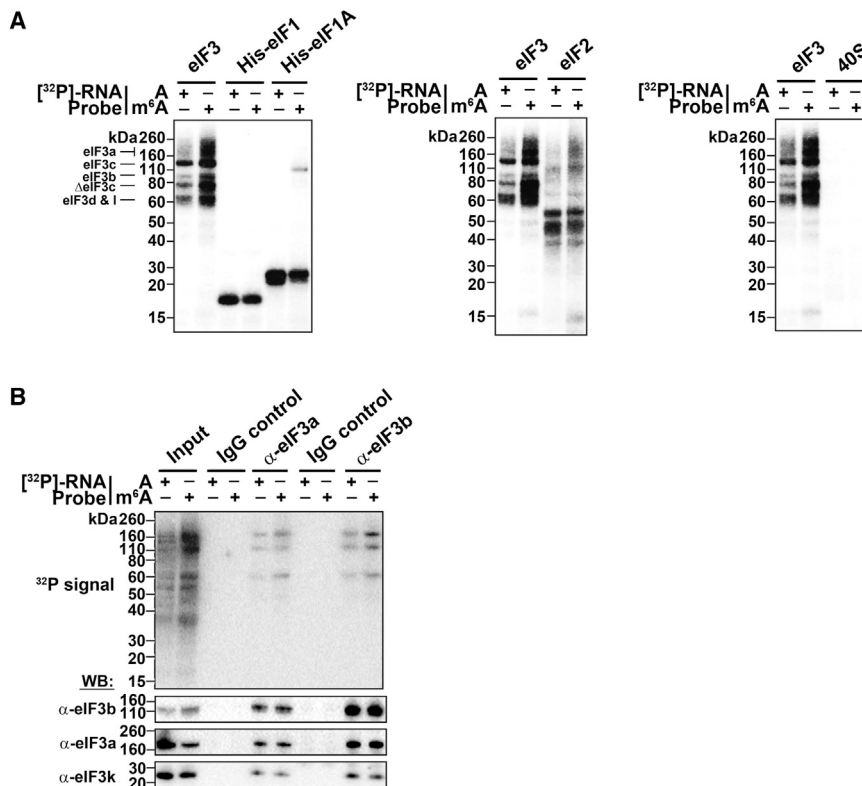
Cell

Figure 4. The 43S Complex Component eIF3 Binds m^{6}A

(A) Indicated proteins/protein complexes were incubated with radiolabeled A- or m^{6}A-containing RNA probes and crosslinked. Unbound RNAs were then removed with RNase I, proteins were separated by SDS-PAGE, and radioactively-labeled RNAs were detected. eIF1, eIF1A, eIF2, and the 40S ribosomal subunit show no preferential crosslinking to methylated RNA. However, eIF3 preparations exhibit strong crosslinking to methylated RNA at bands around 60 kD, 80 kD, and 110–160 kD, which correspond to multiple subunits of the eIF3 complex as indicated.

(B) Crosslinking assays were performed as in (A) using the HeLa cell extracts utilized in in vitro translation assays. The eIF3 complex was immunoprecipitated using antibodies against eIF3a or eIF3b, and proteins containing crosslinked RNA were detected. Both eIF3 antibodies precipitated proteins that preferentially crosslinked to m^{6}A RNA. Immunoprecipitation using rabbit and mouse IgG control antibodies are shown as negative controls. Western blotting for the indicated proteins indicates their enrichment following immunoprecipitation (bottom). The input lanes throughout have 25% of the material loaded for the IP lanes.

See also Figures S2 and S3.

CAG, the m^{6}A-containing probe showed significantly reduced crosslinking to eIF3 (Figure S2E). Thus, efficient eIF3 crosslinking to m^{6}A-containing RNA occurs when the probe contains m^{6}A within its natural sequence context. Furthermore, when we subjected mRNAs that contained a single m^{6}A residue within their 5' UTR to in vitro translation, we found that m^{6}A residues in a GAC context promoted robust cap-independent translation, whereas m^{6}As in a UAC or CAG exhibited markedly reduced translation (Figure S2F). These data indicate that eIF3 preferentially binds to m^{6}A residues in their natural sequence context to promote cap-independent translation.

eIF3 is a large multiprotein complex comprising 13 subunits (a–m) (des Georges et al., 2015) that interacts with mRNA in 48S complexes (Pisarev et al., 2008). UV-crosslinking studies showed that the interaction between eIF3 and RNA occurs at a multisubunit interface (Lee et al., 2015). Similarly, in our crosslinking assays, the m^{6}A-containing probe induced strong labeling of several protein bands, ranging in molecular weight from ∼60 to ∼160 kDa (Figures 4A and S2A–S2E). Particularly strong labeling was observed in the area of ΔeIF3a/eIF3c, ΔeIF3c, and eIF3d/eIF3l (Figures S2A–S2E). These data suggest that m^{6}A-containing RNA may interact with a multisubunit interface within eIF3.

To further explore the binding of m^{6}A-containing RNA to eIF3, we used HeLa cell lysates. Crosslinking using a radioactive m^{6}A-containing RNA probe resulted in the labeling of specific protein bands that were increased relative to the A-containing probe (Figure 4B). Immunoprecipitation of crosslinked extracts using either of two eIF3 subunit-specific antibodies selectively

precipitated these bands, confirming that the increased binding to m^{6}A-containing RNA was mediated by eIF3. Immunoprecipitation with a control antibody recognizing a different initiation-factor-associated protein (ABCF1) did not precipitate these bands (Figures 4B, S3A, and S3B). Thus, these data further suggest that m^{6}A-containing RNA interacts with eIF3.

The m^{6}A-binding protein YTHDF1 interacts with a diverse set of proteins, including eIF3 (Wang et al., 2015). Thus, we considered the possibility that recruitment of eIF3 to m^{6}A-containing RNA in the in vitro translation and crosslinking assays is mediated by a YTH-family m^{6}A-binding protein. However, silver staining of all the initiation factors used in the toeprinting assays failed to show protein bands in the ∼60–64 kD range of these proteins (Figure S2). Additionally, mass spectrometry analysis of the purified eIF3 did not reveal YTH family proteins (Figure S3C) (des Georges et al., 2015). Finally, YTHDF1 was not present in the highly purified eIF3 preparations used in our crosslinking assays, nor were any of the related YTH-domain containing family of m^{6}A binding proteins (Figure S3D) (des Georges et al., 2015). Thus, these data support the idea that eIF3 is able to directly bind m^{6}A.

To determine whether eIF3 binds m^{6}A in cells, we performed PAR-iCLIP to identify zero-distance binding sites of eIF3 in cellular mRNAs. eIF3a-binding sites were primarily localized to 5' UTRs of mRNAs and showed a high degree of overlap with eIF3-binding sites reported previously (Lee et al., 2015) (Figures S4A and S4B).

To determine whether eIF3a binds to sites of m^{6}A in 5' UTRs, we evaluated the overlap of eIF3a-binding sites with m^{6}A

residues mapped at single-nucleotide-resolution in 5′ UTRs (Linder et al., 2015). To test this, we used a permutation-based approach in which eIF3a-binding sites were randomized while preserving the distribution and positional bias of eIF3a PAR-iCLIP tags in 5′ UTRs. Multiple permutations (n > 100) were used, and the statistical significance of overlap between eIF3 PAR-iCLIP sites and m6A residues was evaluated (Supplemental Experimental Procedures). We found a statistically significant overlap between m6A residues and eIF3-binding sites in 5′ UTRs, with 35% of 5′ UTR m6A residues overlapping with eIF3 sites (Figures S4C–S4E). Since single-nucleotide-resolution m6A mapping distinguishes between m6A residues and the m6Am residues that exist as part of the 5′ cap in some mRNAs (Kruse et al., 2011; Linder et al., 2015), we were able to determine that this overlap was specific to m6A residues within 5′ UTRs (Figures S4C and S5A). Taken together, these results support the idea that eIF3 is associated with m6A residues in the 5′ UTRs of cellular mRNAs.

To further test the physiological association of eIF3 and m6A predicted by the PAR-iCLIP analysis, we performed eIF3 protein/RNA immunoprecipitation from HEK293 cells expressing the m6A-demethylating enzyme (Jia et al., 2011), Fto. The abundance of target mRNA 5′ UTRs in the eIF3-bound fraction was then measured using RT-qPCR with primers that amplify the 5′ UTR regions containing the m6A residue. mRNAs that contain a high stoichiometry m6A site within their 5′ UTR (Meyer et al., 2012) were substantially depleted in the eIF3-bound fraction following Fto overexpression (Figure 5B). In contrast, eIF3 immunoprecipitation of a control mRNA deficient in 5′ UTR m6A (Meyer et al., 2012) was unaffected by Fto overexpression (Figure 5B). Taken together, these data support the idea that eIF3 interacts with mRNAs in an m6A-dependent manner in cells.

m6A within the 5′ UTR Promotes Translation of Cellular mRNAs

To address whether mRNAs that contain 5′ UTR m6A residues possess enhanced translation in cells, we examined ribosome profiling-based measurements of mRNA translation efficiency (TE) in HeLa cells depleted of the m6A methyltransferase enzyme, METTL3, which results in depletion of all m6A residues in cells (Wang et al., 2015). We examined the TE of mRNAs based on the location of their m6A residues identified by single-nucleotide-resolution m6A mapping (Linder et al., 2015). Compared to mRNAs that lack m6A, we found that transcripts that contain m6A residues within the coding sequence or 3′ UTR show no significant change in TE in METTL3-depleted cells (Figures 6A and 6B). Similarly, mRNAs that contain m6A residues near the stop codon do not show reduced translation in METTL3-depleted cells. However, mRNAs containing 5′ UTR m6A residues showed a large reduction in TE following METTL3 depletion, suggesting a preferential role for 5′ UTR m6A in promoting mRNA translation (Figures 6A, 6B, and S5B). Residual translation may reflect ongoing cap-dependent translation in METTL3-deficient cells. The translation of mRNAs containing 5′ UTR m6A residues was not suppressed in cells depleted of YTHDF1 (Figure S5B), which is consistent with the idea that 5′ UTR m6A promotes translation through eIF3. Taken together, these data

suggest that m6A residues in the 5′ UTR enhance the translation of mRNAs in cells.

Heat-Shock-Induced Translation of Hsp70 Is Mediated by 5′ UTR m6A

We next sought to investigate the role of m6A in promoting cap-independent translation in cells. Since cellular translation involves both cap-dependent and cap-independent mechanisms, we took advantage of heat shock, which induces a stress response that suppresses most cap-dependent translation (Holcik and Sonenberg, 2005). Heat-shock protein 70 (HSP70) is a stress response mRNA known to undergo increased transcription and cap-independent translation following heat shock (Lindquist and Craig, 1988). Previous studies demonstrated that HSP70 contains an m6A site within its 5′ UTR (Schwartz et al., 2014) and that methylation of the HSP70 5′ UTR is increased following heat shock (Dominissini et al., 2012). However, the role of m6A in cap-independent translation of HSP70 is not understood.

To test the effect of m6A in HSP70 translation, we utilized altered expression of Fto to influence m6A levels within the Hsp70 5′ UTR. Knockdown of Fto resulted in increased m6A levels in Hsp70 mRNA in heat-shocked cells (Figure S6A). Conversely, overexpressing Fto in heat-shocked cells reduced the level of m6A in Hsp70 mRNA by 29% relative to heat-shocked cells overexpressing GFP (Figure S6A). To determine whether altered m6A levels in the Hsp70 5′ UTR influence heat shock-induced Hsp70 translation, we used mouse embryonic fibroblasts (MEFs), which exhibit low Hsp70 levels prior to heat shock (Sun et al., 2011). In MEF cells stably expressing control shRNA, Hsp70 protein was readily detected 4 and 6 hr after heat shock. However, in MEF cells stably expressing Fto-specific shRNA to increase m6A levels, Hsp70 protein expression was significantly higher at both 4 and 6 hr after heat shock (Figure 6C). This effect was not due to increased levels of Hsp70 mRNA (Figure S6B). Furthermore, knockdown of Fto caused a significant increase in the fraction of polysome-bound Hsp70 mRNA (Figure 6D), suggesting that the increased levels of Hsp70 protein seen after heat shock reflect increased translation of Hsp70 mRNA in Fto knockdown cells.

Consistent with the effects of Fto knockdown on Hsp70 levels, Fto overexpression caused significantly reduced Hsp70 protein production 4 and 6 hr after heat shock (Figure 6E). This effect was not due to reduced Hsp70 transcript levels (Figure S6B). In addition, Hsp70 mRNA was significantly reduced in the polysome fractions of Fto-overexpressing cells compared to GFP-expressing cells, confirming that the Fto-mediated reduction in Hsp70 protein levels was due to reduced Hsp70 translation (Figure 6F). These data suggest that the loss of m6A in Hsp70 mRNA results in reduced translation efficiency following heat shock.

Transcriptome-wide Redistribution of m6A following Cellular Stress

We next sought to further understand the importance of 5′ UTR m6A residues in response to cellular stress. Based on our findings with Hsp70 mRNA, we considered the possibility that heat shock may alter the transcriptome-wide distribution of m6A. Under basal conditions, most m6A residues are located in
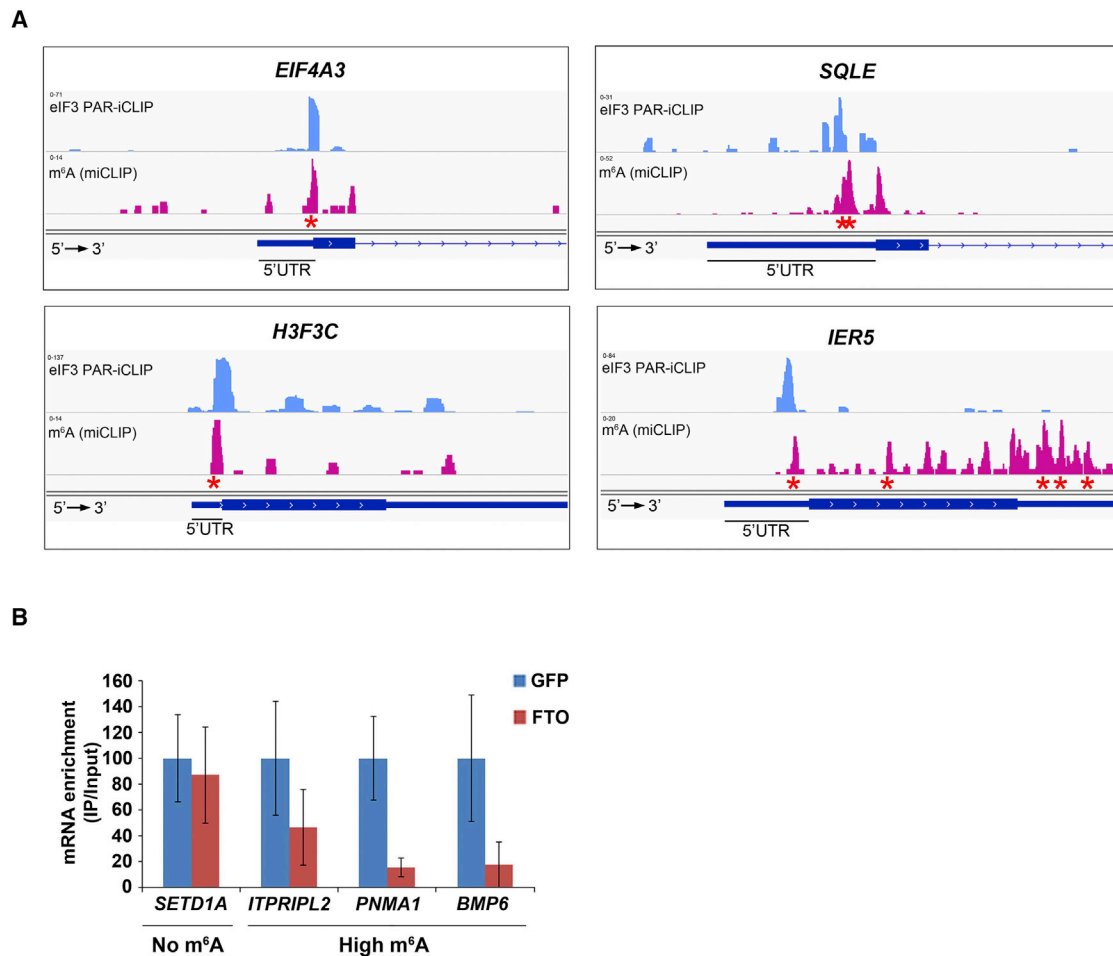
Cell

Figure 5. eIF3 Binding Sites within Cellular mRNAs Localize to Sites of m6A Residues within the 5′ UTR

(A) Shown are read clusters from both eIF3 PAR-iCLIP (light blue) and single-nucleotide-resolution m6A mapping (Linder et al., 2015) (miCLIP; red) for four representative mRNAs (EIF4A3, H3F3C, SQLE, and IER5). eIF3a PAR-iCLIP read clusters exhibit highly specific overlap with m6A mapping clusters at internal positions within 5′ UTRs. This co-localization is specific to 5′ UTRs, as mRNAs that contain multiple m6A residues in the CDS or 3′ UTR fail to show eIF3a binding at these sites (exemplified by IER5). Red asterisks indicate the location of individual m6A sites identified at single-nucleotide resolution.

(B) eIF3 binds to the 5′ UTR of cellular mRNAs in an m6A-dependent manner. HEK293 cells were transfected with GFP- or Fto-overexpression plasmids, and eIF3 immunoprecipitation was performed to isolate eIF3-bound mRNAs. Bound mRNAs were quantified by RT-qPCR using 5′ UTR-specific primers. 5′ UTRs of mRNAs that contain high levels of m6A exhibited reduced binding to eIF3 after overexpression of Fto. 5′ UTRs that do not contain m6A exhibited no change in eIF3 binding following Fto overexpression (n = 3; mean ± SEM).

See also Figures S4 and S5.

mRNAs near the stop codon, with markedly fewer m6A residues in 5′ UTRs. To determine if cellular stress alters the characteristic distribution of m6A, we mapped m6A residues using miCLIP, a method for single-nucleotide resolution detection of m6A sites (Linder et al., 2015). Remarkably, the metagene analysis showed a marked enrichment of m6A in the 5′ UTR in heat-shocked cells compared to control cells (Figure S6C).

To further examine this phenomenon, we analyzed existing transcriptome-wide m6A mapping datasets that were performed in stressed cells and control cells. These include HepG2 cells treated with UV, interferon-γ, and heat shock (Dominissini et al., 2012). Metagene analyses showed prominent increases in the level of 5′ UTR m6A in both the UV-treated and heat-shocked cells (Figure S6D). The number of m6A sites in the

3′ UTR was relatively unaffected following heat shock or UV compared to control (n = 4,538, 4,533, or 3,171, respectively), whereas the number of m6A sites in the 5′ UTR was markedly increased in heat shock and UV relative to control (n = 1,501, 1,212, or 326, respectively) (Table S1). Notably, interferon-γ treatment did not alter the m6A metagene profile (Figure S6D), indicating that the induction of 5′ UTR m6A is not a nonspecific stress response but instead is linked to specific forms of cellular stress.

Intriguingly, both heat shock and UV caused increased 5′ UTR methylation in mRNAs that belong to common functional pathways, including phosphorylation and cell-cycle regulation (Table S1). Collectively, our results indicate that activation of some stress-response pathways causes a global reshaping of

Cell

the cellular mRNA methylome and suggest that increased 5′ UTR methylation may be a general component of the response to select cellular stresses. Future studies will be important for understanding how stress pathways increase m6A within the 5′ UTR of mRNAs and reshape the RNA methylome. Furthermore, it will be important to analyze how diverse stress response pathways utilize these upregulated 5′ UTR m6A residues to mediate translational responses.

DISCUSSION

Eukaryotic mRNAs can be translated in both cap-dependent and cap-independent modes, although the mechanisms of translation initiation that do not require the 5′ cap and eIF4E have been poorly understood. Our results show that m6A residues within the 5′ UTR can act as an m6A-induced ribosome engagement site (MIRES), which promotes cap-independent translation of mRNA. We find that a single m6A in the 5′ UTR of mRNAs is sufficient to promote MIRES activity in cell-free extracts, whereas m6A residues outside the 5′ UTR fail to show this effect. The significance of 5′ UTR m6A residues is further seen in both ribosome profiling datasets and in individual cellular mRNAs in conditions where cap-dependent translation is suppressed. These results point to selective recognition of 5′ UTR m6A as a mechanism for mRNAs to bypass the cap requirement for translation and suggest a potential role for this class of m6A residues in mediating translational responses induced in diverse cellular stress states.

A role for m6A in promoting translation initiation is supported by our finding that METTL3 depletion leads to a large reduction in translation efficiency of mRNAs containing 5′ UTR m6A residues compared to mRNAs that contain m6As elsewhere. Although cap-independent translation of cellular mRNAs may also be mediated by m6A-independent pathways, including direct recruitment of ribosomes to internal 5′ sequence or structural elements (Xue et al., 2015), our studies raise the intriguing possibility that an eIF4E-independent mode of translation initiation can be switched on or off by reversible methylation of adenosine residues in the 5′ UTR of mRNAs.

Our studies show that cap-independent translation mediated by m6A requires a novel m6A reader, eIF3. We find that many eIF3-binding sites in the transcriptome overlap with m6A sites in 5′ UTRs. The identification of eIF3 as an m6A reader was originally suggested by the finding that the 48S complex can be assembled on m6A-containing RNA using only eIF1, eIF1A, eIF2, eIF3, and the 40S subunit. Of these components, eIF3 shows selective interaction with m6A both in vitro and in cells. By binding eIF3, 5′ UTR m6A residues can stimulate translation initiation by directly recruiting the 43S preinitiation complex to the 5′ UTR of mRNAs.

m6A has diverse effects on mRNAs, including mRNA destabilization and translational enhancement, although these effects are predominantly attributed to m6A near stop codons or in 3′ UTRs (Wang et al., 2014a, 2015). In the case of m6A near stop codons or in 3′ UTRs, translational enhancement is mediated by YTHDF1, which binds to select transcripts at m6A sites in their 3′ UTRs (Wang et al., 2015). YTHDF1 binds numerous proteins, including eIF3 and other ribosome-associated proteins, which are proposed

to be recruited to 3′ UTRs to influence cap-dependent translation initiation (Wang et al., 2015). This is in contrast to the mechanism of 5′ UTR m6A, which directly recruits eIF3 and assembles translation initiation complexes in the 5′ UTR without cap-binding proteins. Our analysis of ribosome profiling data from YTHDF1-depleted cells further indicates that 5′ UTR m6A residues promote translation through a YTHDF1-independent mechanism. Thus, m6A exhibits markedly distinct effects on mRNA based on its location in transcripts.

A long-standing question is the mechanism by which select cellular mRNAs undergo cap-independent translation during conditions where cap-dependent translation is suppressed (Holcik and Sonenberg, 2005). A prevailing hypothesis has been that these mRNAs contain cellular IRESs that promote cap-independent translation (Komar and Hatzoglou, 2011). However, putative cellular IRESs often lack the complex structural elements seen in viral IRESs (Hellen and Sarnow, 2001). As a result of this discrepancy, and because of flaws inherent to many assays that test cellular IRES function, the evidence for and against cellular IRESs is a frequent topic of debate (Gilbert, 2010; Kozak, 2005). Given the prevalence of m6A within 5′ UTRs, their translation-promoting activity represents an additional or perhaps alternative mechanism for mediating cap-independent translation.

The importance of 5′ UTR m6A residues is supported by their selective upregulation in response to specific forms of stress. This m6A stress response points to the importance of this subset of m6A residues, which our results show are linked to cap-independent translation. Notably, other forms of stress regulate translation through the integrated stress response (Ron, 2002). It will be important to determine if 5′ UTR m6A-mediated translation is an alternative mechanism to orchestrate translational responses to stress.

EXPERIMENTAL PROCEDURES

In Vitro Translation

In vitro translation assays were performed using HeLa cell extracts (One-Step Human IVT Kit, Thermo Scientific). Equal amounts of RNA were used for each reaction (100 ng RNA per reaction, ~30 nM per reaction), and all reactions within each experiment were performed in equal volumes. Multiple different batches of HeLa extracts and mRNA preparations were used to ensure that the translation-promoting effect of m6A is not due to a specific lot of extract or batch of synthesized mRNA. However, this also contributes to inter-experiment variability. Reactions were performed at 30°C for 30 min and were stopped by the addition of 200 μM cycloheximide and placed on ice. 1 μl of each reaction was then used for luminescence analysis (see below). The remaining reaction volume was used for RNA isolation with TRIzol (Invitrogen) or QIAGEN RNeasy kits according to the manufacturer's instructions. cDNA synthesis was then performed using Superscript III reverse transcriptase (Invitrogen) and random hexamers. Following treatment with RNase H, cDNA was then used for RT-qPCR analysis to ensure that differences in mRNA levels across samples did not account for the observed changes in protein production. Statistical analysis of luciferase activity measurements (below) was performed using Student's t test and a p value threshold of 0.01.

Luciferase Activity Measurements

Luciferase expression was measured using the One-Glo luciferase assay kit (Promega) according to the manufacturer's instructions. Luminescence measurements were performed on a Molecular Devices Spectramax L microplate reader using the SoftMax Pro software program.
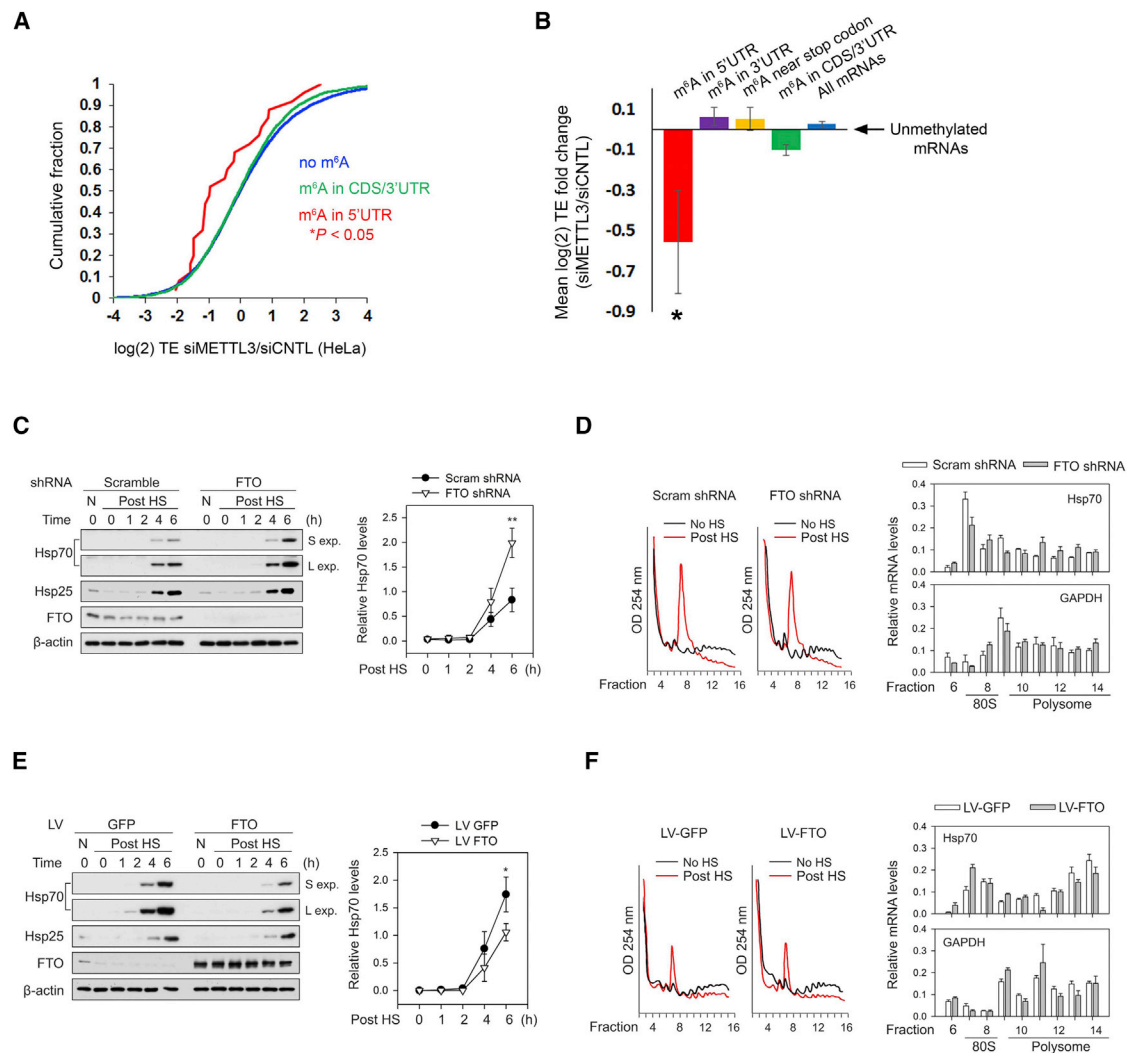
Cell

Figure 6. m⁶A Mediates Stress-Induced Translation of *Hsp70*

(A) Depletion of the m⁶A methyltransferase, METTL3, decreases the TE of mRNAs with 5′ UTR m⁶A. Ribosome profiling data from HeLa cells expressing METTL3 or control siRNAs (Wang et al., 2015) were used to determine changes in TE for various classes of mRNAs defined by single-nucleotide-resolution m⁶A mapping. Compared to nonmethylated mRNAs (blue), transcripts with m⁶A residues in the coding sequence (CDS) or 3′ UTR (green) exhibit only a marginal decrease in TE. However, mRNAs containing m⁶A within the 5′ UTR (red) show a large reduction in TE. p values were calculated using the Mann-Whitney test.

(B) TEs of various classes of m⁶A-containing mRNAs were analyzed using ribosome profiling datasets from HeLa cells as described in (A). Shown are the mean fold changes in TE (siMETTL3/siControl) for mRNAs with m⁶A residues only in the 5′ UTR (red), within the 3′ UTR (purple), within 50 nt of the stop codon (yellow), within the CDS and/or 3′ UTR (green), or in all mRNAs (blue), as defined by single-nucleotide-resolution m⁶A mapping. mRNAs with 5′ UTR m⁶A residues exhibit a dramatic reduction in TE after METTL3 depletion, whereas transcripts with m⁶As in other regions fail to show this effect. All mean fold change TE values were computed after background subtraction of the mean fold change computed from all nonmethylated control mRNAs, as indicated by the arrow (mean ± SEM; *p < 0.05).

(C) *Fto* knockdown increases heat-shock-induced translation of *Hsp70*. MEF cells stably expressing either *Fto* shRNA or scramble shRNA were subjected to heat shock stress. Cell lysates were collected at various times post-heat shock ("Post HS") and then used for western blot analysis with the indicated antibodies. *Fto* knockdown increased the levels of stress-induced Hsp70 protein compared to control shRNA ("S exp" = short exposure; "L exp" = long exposure). Levels of Hsp25, another heat shock-induced protein, were unaffected by *Fto* knockdown. Right panel shows quantification of Hsp70 levels normalized to β-actin (n = 3; mean ± SEM; **p < 0.1).

(D) MEFs stably expressing control or *Fto* shRNA were subjected to heat shock stress as in (C). Polysome fractions were separated using sucrose gradient fractionation (left panels) followed by RT-qPCR for *Hsp70* (top right panel) and *Gapdh* (bottom right panel) in each fraction. *Hsp70* levels are increased in polysome fractions following *Fto* knockdown, whereas the distribution of *Gapdh* is unchanged (n = 3; mean ± SEM; *Hsp70*: p = 0.0007, two-way ANOVA; *Gapdh*: p = 0.3722, two-way ANOVA considering the entire range of time points).

(E) MEF cells were infected with either GFP or Fto lentivirus and subjected to heat shock stress. Cell lysates were collected at various times post-heat shock and then used for western blot analysis with the indicated antibodies. Fto overexpression decreased the levels of heat-shock-induced Hsp70 protein compared to GFP overexpression. Levels of Hsp25 were unaffected by Fto overexpression. Right panel shows quantification of Hsp70 levels normalized to β-actin (n = 3; Mean ± SEM; *p < 0.5).

(legend continued on next page)

Cell

eIF3a PAR-iCLIP

eIF3a PAR-iCLIP was performed using HEK293T cells as described previously (Huppertz et al., 2014) with some adjustments. 10 million cells were incubated with 100 mM 4SU for 8 hr. Media was then discarded, and cells were placed on ice and irradiated with 365 nm UV light using a Stratalinker UV crosslinker (Stratagene) with 150 mJ/cm^2. Cells were scraped in ice-cold 1× PBS and collected by centrifugation at 200 x g for 10 min at 4°C. Cell pellets were suspended in 200 μl of 1% SDS, 10 mM DTT, and 1× protease inhibitors (cOmplete mini EDTA-free, Roche). The lysate was then passed through an 18G needle 10 times to improve cell lysis and shearing of DNA. SDS was neutralized by diluting the lysate to 2 ml using RIPA buffer without SDS. The remainder of the protocol was performed as described (Huppertz et al., 2014) using rabbit anti-eIF3a (Abcam).

Additional methods are detailed in the Supplemental Experimental Procedures section.

ACCESSION NUMBERS

Sequencing data have been deposited to NCBI's GEO under accession number GEO: GSE73405.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.012.

AUTHOR CONTRIBUTIONS

K.D.M., T.V.P., S.-B.Q., and S.R.J. designed the experiments and analyzed data. J.Z. and S.-B.Q. performed experiments related to Fto and Hsp70. K.D.M. and S.R.J. wrote the manuscript. K.D.M., D.P.P., J.Z., A.Z., and M.A.S. performed the experiments, collected data, and prepared the figures. O.E. and D.P.P. performed computational analysis of sequencing data. All authors commented and made edits to the manuscript.

REFERENCES

De Gregorio, E., Preiss, T., and Hentze, M.W. (1998). Translational activation of uncapped mRNAs by the central part of human eIF4G is 5′ end-dependent. RNA *4*, 828–836.

des Georges, A., Dhote, V., Kuhn, L., Hellen, C.U., Pestova, T.V., Frank, J., and Hashem, Y. (2015). Structure of mammalian eIF3 in the context of the 43S pre-initiation complex. Nature *525*, 491–495.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206.

Gilbert, W.V. (2010). Alternative ways to think about cellular internal ribosome entry. J. Biol. Chem. *285*, 29033–29038.

Gingras, A.C., Raught, B., and Sonenberg, N. (1999). eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. Annu. Rev. Biochem. *68*, 913–963.

Hellen, C.U., and Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. Genes Dev. *15*, 1593–1612.

Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. Nat. Rev. Mol. Cell Biol. *6*, 318–327.

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., König, J., and Ule, J. (2014). iCLIP: protein-RNA interactions at nucleotide resolution. Methods *65*, 274–287.

Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. *11*, 113–127.

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.G., et al. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat. Chem. Biol. *7*, 885–887.

Komar, A.A., and Hatzoglou, M. (2011). Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. Cell Cycle *10*, 229–240.

Kozak, M. (2005). A second look at cellular mRNA sequences said to function as internal ribosome entry sites. Nucleic Acids Res. *33*, 6593–6602.

Kruse, S., Zhong, S., Bodi, Z., Button, J., Alcocer, M.J., Hayes, C.J., and Fray, R. (2011). A novel synthesis and detection method for cap-associated adenosine modifications in mouse mRNA. Sci. Rep. *1*, 126.

Lee, A.S., Kranzusch, P.J., and Cate, J.H. (2015). eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. Nature *522*, 111–114.

Linder, B., Grozhik, A.V., Olarerin-George, A.O., Meydan, C., Mason, C.E., and Jaffrey, S.R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat. Methods *12*, 767–772.

Lindquist, S., and Craig, E.A. (1988). The heat-shock proteins. Annu. Rev. Genet. *22*, 631–677.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell *149*, 1635–1646.

Mikami, S., Masutani, M., Sonenberg, N., Yokoyama, S., and Imataka, H. (2006). An efficient mammalian cell-free translation system supplemented with translation factors. Protein Expr. Purif. *46*, 348–357.

Pestova, T.V., and Hellen, C.U. (2003). Translation elongation after assembly of ribosomes on the Cricket paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. Genes Dev. *17*, 181–186.

Pestova, T.V., and Kolupaeva, V.G. (2002). The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. Genes Dev. *16*, 2906–2922.

Pisarev, A.V., Kolupaeva, V.G., Yusupov, M.M., Hellen, C.U., and Pestova, T.V. (2008). Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. EMBO J. *27*, 1609–1621.

Ray, P.S., Grover, R., and Das, S. (2006). Two internal ribosome entry sites mediate the translation of p53 isoforms. EMBO Rep. *7*, 404–410.

Ron, D. (2002). Translational control in the endoplasmic reticulum stress response. J. Clin. Invest. *110*, 1383–1388.

(F) MEFs with or without Fto overexpression were subjected to heat shock stress as in (E). Polysome fractions were separated using sucrose gradient fractionation (left panels) followed by RT-qPCR of *Hsp70* (top right panel) and *Gapdh* (bottom right panel) in each fraction. *Hsp70* levels are decreased in polysome fractions following Fto overexpression, whereas the distribution of *Gapdh* is unchanged (n = 3; mean ± SEM; *Hsp70*: $p < 0.0001$, two-way ANOVA; *Gapdh*: $p = 0.1910$, two-way ANOVA considering the entire range of time points).
See also Figures S5 and S6 and Table S1.

Cell

Roost, C., Lynch, S.R., Batista, P.J., Qu, K., Chang, H.Y., and Kool, E.T. (2015). Correction to "Structure and Thermodynamics of N(6)-Methyladenosine in RNA: A Spring-Loaded Base Modification". J. Am. Chem. Soc. *137*, 8308.

Schwartz, S., Mumbach, M.R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G.G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D., et al. (2014). Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5′ sites. Cell Rep. *8*, 284–296.

Stoneley, M., and Willis, A.E. (2004). Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. Oncogene *23*, 3200–3207.

Sun, J., Conn, C.S., Han, Y., Yeung, V., and Qian, S.B. (2011). PI3K-mTORC1 attenuates stress response by inhibiting cap-independent Hsp70 translation. J. Biol. Chem. *286*, 6791–6800.

Terenin, I.M., Andreev, D.E., Dmitriev, S.E., and Shatsky, I.N. (2013). A novel mechanism of eukaryotic translation initiation that is neither m7G-cap-, nor IRES-dependent. Nucleic Acids Res. *41*, 1807–1816.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014a). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Wang, Y., Li, Y., Toth, J.I., Petroski, M.D., Zhang, Z., and Zhao, J.C. (2014b). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat. Cell Biol. *16*, 191–198.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell *161*, 1388–1399.

Xue, S., Tian, S., Fujii, K., Kladwang, W., Das, R., and Barna, M. (2015). RNA regulons in Hox 5′ UTRs confer ribosome specificity to gene regulation. Nature *517*, 33–38.

Cell

The Molecular Taxonomy of Primary Prostate Cancer

Graphical Abstract
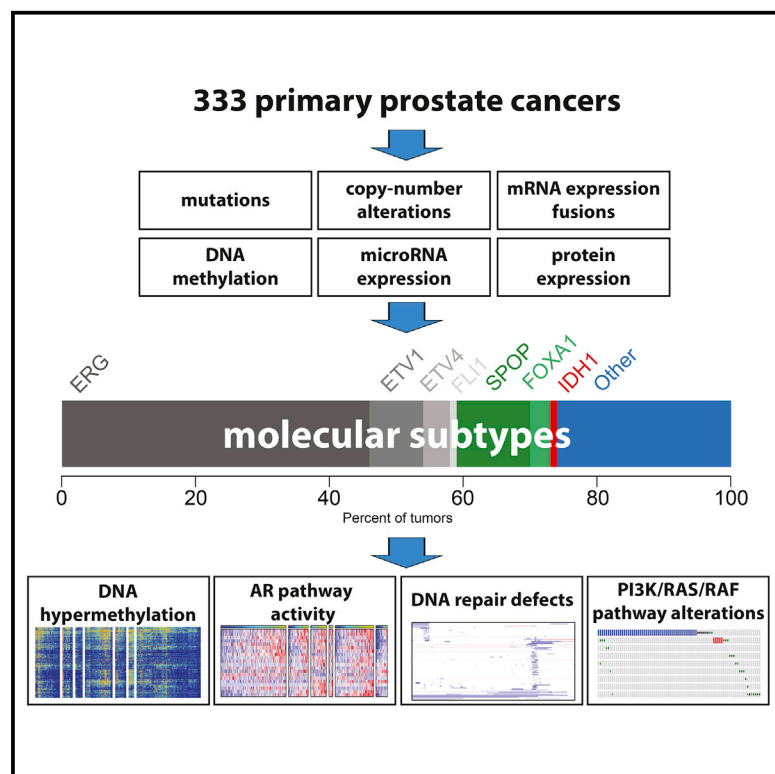
Authors

The Cancer Genome Atlas Research Network

Correspondence

schultz@cbio.mskcc.org (N.S.),
massimo_loda@dfci.harvard.edu (M.L.),
sander.research@gmail.com (C.S.)

In Brief

Molecular analysis of 333 primary prostate carcinomas reveals substantial heterogeneity and major subtypes among patients, as well as potentially actionable lesions valuable for clinical management of the disease.

Highlights

- Comprehensive molecular analysis of 333 primary prostate carcinomas

- Seven subtypes defined by ETS fusions or mutations in *SPOP*, *FOXA1*, and *IDH1*

- Substantial epigenetic heterogeneity, including a hypermethylated *IDH1* mutant subset

- Presumed actionable lesions in the PI3K, MAPK, and DNA repair pathways

CrossMark

CellPress

The Molecular Taxonomy of Primary Prostate Cancer

The Cancer Genome Atlas Research Network[1,*]
[1]The Cancer Genome Atlas Program Office, National Cancer Institute at NIH, 31 Center Drive, Building 31, Suite 3A20, Bethesda, MD 20892, USA
*Correspondence: schultz@cbio.mskcc.org (N.S.), massimo_loda@dfci.harvard.edu (M.L.), sander.research@gmail.com (C.S.)
http://dx.doi.org/10.1016/j.cell.2015.10.025

SUMMARY

There is substantial heterogeneity among primary prostate cancers, evident in the spectrum of molecular abnormalities and its variable clinical course. As part of The Cancer Genome Atlas (TCGA), we present a comprehensive molecular analysis of 333 primary prostate carcinomas. Our results revealed a molecular taxonomy in which 74% of these tumors fell into one of seven subtypes defined by specific gene fusions (*ERG*, *ETV1/4*, and *FLI1*) or mutations (*SPOP*, *FOXA1*, and *IDH1*). Epigenetic profiles showed substantial heterogeneity, including an *IDH1* mutant subset with a methylator phenotype. Androgen receptor (AR) activity varied widely and in a subtype-specific manner, with *SPOP* and *FOXA1* mutant tumors having the highest levels of AR-induced transcripts. 25% of the prostate cancers had a presumed actionable lesion in the PI3K or MAPK signaling pathways, and DNA repair genes were inactivated in 19%. Our analysis reveals molecular heterogeneity among primary prostate cancers, as well as potentially actionable molecular defects.

INTRODUCTION

Prostate cancer is the second most common cancer in men and the fourth most common tumor type worldwide (Ferlay et al., 2013). It is estimated that, in 2015, prostate cancer will be diagnosed in 220,800 men in the United States alone and that 27,540 will die of their disease (Siegel et al., 2015). Multiple genetic and demographic factors, including age, family history, genetic susceptibility, and race, contribute to the high incidence of prostate cancer (Al Olama et al., 2014).

In the current era of prostate-specific antigen (PSA) screening, nearly 90% of prostate cancers are clinically localized at the time of their diagnosis (Penney et al., 2013). The clinical behavior of localized prostate cancer is highly variable—while some men will have aggressive cancer leading to metastasis and death from the disease, many others will have indolent cancers that are cured with initial therapy or may be safely observed. Multiple risk stratification systems have been developed, combining the best currently available clinical and pathological parameters (such as Gleason score, PSA levels, and clinical and pathological staging); however, these tools still do not adequately predict outcome (Cooperberg et al., 2009; D'Amico et al., 1998; Kattan et al., 1998). Further risk stratification using molecular features could potentially help distinguish indolent from aggressive prostate cancer.

Molecular and genetic profiles are increasingly being used to subtype cancers of all types and to guide selection of more precisely targeted therapeutic interventions. Several recent studies have explored the molecular basis of primary prostate cancer and have identified multiple recurrent genomic alterations that include mutations, DNA copy-number changes, rearrangements, and gene fusions (Baca et al., 2013; Barbieri et al., 2012; Berger et al., 2011, Lapointe et al., 2007; Pflueger et al., 2011; Taylor et al., 2010; Tomlins et al., 2007; Wang et al., 2011). The most common alterations in prostate cancer genomes are fusions of androgen-regulated promoters with *ERG* and other members of the E26 transformation-specific (ETS) family of transcription factors. In particular, the *TMPRSS2-ERG* fusion is the most common molecular alteration in prostate cancer (Tomlins et al., 2005), being found in between 40% and 50% of prostate tumor foci, translating to more than 100,000 cases annually in the United States (Tomlins et al., 2009). Nevertheless, among treated prostate cancers, and despite extensive study, affected individuals with fusion-bearing tumors do not appear to have a significantly different prognosis following prostatectomy than those without (Gopalan et al., 2009; Pettersson et al., 2012). Prostate cancers also have varying degrees of DNA copy-number alteration; indolent and low-Gleason tumors have few alterations, whereas more aggressive primary and metastatic tumors have extensive burdens of copy-number alteration genome wide (Taylor et al., 2010; Hieronymus et al., 2014; Lalonde et al., 2014). In contrast, somatic point mutations are less common in prostate cancer than in most other solid tumors. The most frequently mutated genes in primary prostate cancers are *SPOP*, *TP53*, *FOXA1*, and *PTEN* (Barbieri et al., 2012). Only recently has the spectrum of epigenetic changes in prostate cancer genomes been explored (Börno et al., 2012; Friedlander et al., 2012; Kim et al., 2011; Kobayashi et al., 2011; Mahapatra et al., 2012).

Importantly, no studies have comprehensively integrated diverse omics data types to assess the robustness of previously defined subtypes and potentially prognostic alterations. Here, to gain further insight into the molecular-genetic heterogeneity of primary prostate cancer and to establish a molecular taxonomy of the disease for future diagnostic, prognostic, and therapeutic stratification, the TCGA Network has comprehensively characterized 333 primary prostate cancers using seven genomic platforms. This analysis reveals novel molecular features that provide a better understanding of this disease and suggest potential therapeutic strategies.

Table 1. Cohort Characteristics

Clinical Feature	
Age	61 (43–76)
Pre-operative PSA	7.4 (1.6–87.0)
Gleason Score	
3+3	65
3+4	102
4+3	78
≥ 8	88
Tumor Cellularity (pathology)	
<20%	7
21–40%	40
41–60%	84
61–80%	115
81–100%	87
Pathologic Stage	
pT2a/b	18
pT2c	111
pT3a	110
pT3b	82
pT4	6
PSA Recurrence	
Yes	33
No[a]	248
Not available	47
Margin Status	
Positive	69
Negative	193
Not available	71
Ethnicity	
Caucasian	270
African descent	43
Asian	8
Not available	12

[a]Either no evidence of recurrence or insufficient follow-up.

RESULTS

Cohort and Platforms

The cohort of primary prostate cancers analyzed resulted from extensive pathologic, analytical, and quality control review, yielding 333 tumors from 425 available cases. Images of frozen tissue were evaluated by multiple expert genitourinary pathologists, and cases were excluded if no tumor cells were identifiable in the sample or if there was evidence of significant RNA degradation (Figure S1; Supplemental Experimental Procedures). For the subset of cases reviewed by two pathologists, tumor cellularity estimates were within 20% of each other in 71% of cases. In total, 78% of Gleason scores were concordant within one grade of the secondary pattern (Supplemental Experimental Procedures). Moreover, due to the challenge of acquiring primary

prostate cancer specimens of high tumor cellularity, we also performed a multi-platform analysis of tumor content, estimating tumor purity with analytical approaches utilizing both DNA (Carter et al., 2012; Prandi et al., 2014) and RNA (Quon et al., 2013; Ahn et al., 2013) sequencing data. The molecular and pathologic estimates are presented in Table S1A and Figure S1. The clinical and pathological characteristics of the final cohort are presented in Table 1. The average follow-up time following radical prostatectomy was just under 2 years, which precluded outcomes analysis due to the long natural history of primary prostate cancer.

We characterized isolated biomolecules from these 333 tumor samples using four platforms: whole-exome sequencing for somatic mutations, array-based methods for profiling both somatic copy-number changes and DNA methylation, and mRNA sequencing. We also performed microRNA (miRNA) sequencing on 330 of these samples, reverse-phase protein array (RPPA) on 152 samples, and low-pass and high-pass whole-genome sequencing (WGS) on 100 and 19 tumor/normal pairs, respectively (Supplemental Experimental Procedures). For 19 samples, non-malignant adjacent prostate samples were also examined for DNA methylation and RNA/miRNA expression analyses.

The Molecular Taxonomy of Primary Prostate Cancer

Previous studies indicate that many genomically distinct subsets of prostate cancer exist. These are driven in some cases by frequent events, such as androgen-regulated fusions of *ERG* and other ETS family members, or recurrent *SPOP* mutations and, in other cases, by less common genomic aberrations. Given the comprehensive nature of our data, we sought to unify these disparate findings to establish a molecular taxonomy of primary disease that integrates results from somatic mutations, gene fusions, somatic copy-number alterations (SCNA), gene expression, and DNA methylation. We first performed unsupervised clustering of data from each molecular platform, as well as integrative clustering using iCluster (Shen et al., 2009) (Figures S2, S3, S4, S5, S6, and S7). These analyses uncovered both known and novel associations, with 74% of all tumors being assignable to one of seven molecular classes based on distinct oncogenic drivers: fusions involving (1) *ERG*, (2) *ETV1*, (3) *ETV4*, or (4) *FLI1* (46%, 8%, 4%, and 1%, respectively); mutations in (5) *SPOP* or (6) *FOXA1*; or (7) *IDH1* mutations (11%, 3%, and 1%, respectively) (Figures 1 and S2 and Table S1A).

In total, 53% of tumors were found to have ETS family gene fusions (*ERG*, *ETV1*, *ETV4*, and *FLI1*) after analysis with two complementary algorithms (Sboner et al., 2010; Wang et al., 2010) (see the Experimental Procedures). While *TMPRSS2* was the most frequent fusion partner in all ETS fusions, we identified fusions with other previously described androgen-regulated 5′ partner genes, including *SLC45A3* and *NDRG1* (Table S1E). We also identified several tumors that overexpressed full-length ETS transcripts that were mutually exclusive with ETS fusions (12 *ETV1* high tumors, 6 *ETV4*, and 2 *FLI1*) (Table S1E). ETS overexpression in these cases could possibly be mediated via epigenetic mechanisms or cryptic translocations of the entire gene locus to a transcriptionally active neighborhood. In the one case with elevated ETV1 full-length expression studied by whole-genome sequencing, we identified a cryptic genomic
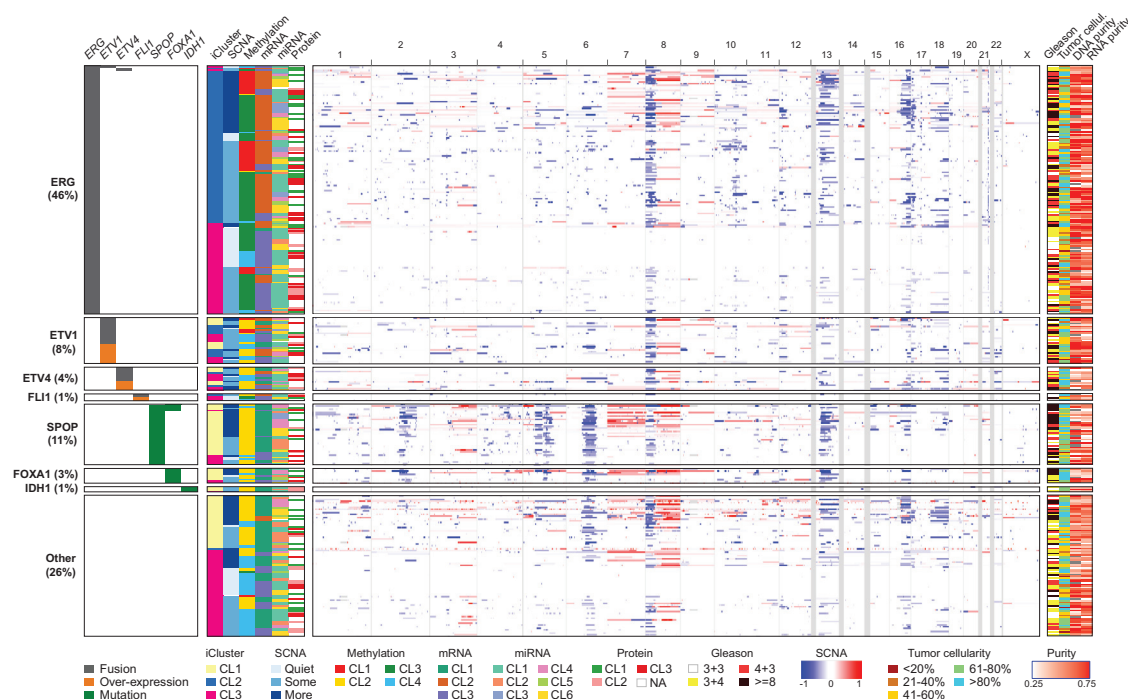
Figure 1. The Molecular Taxonomy of Primary Prostate Cancer

Comprehensive molecular profiling of 333 primary prostate cancer samples revealed seven genomically distinct subtypes, defined (top to bottom) by *ERG* fusions (46%), *ETV1/ETV4/FLI1* fusions or overexpression (8%, 4%, 1%, respectively), or by *SPOP* (11%), *FOXA1* (3%), and *IDH1* (1%) mutations. A subset of these subtypes was correlated with clusters computationally derived from the individual characterization platforms (somatic copy-number alterations, methylation, mRNA, microRNA, and protein levels from reverse phase protein arrays). The heatmap shows DNA copy-number for all cases, with chromosomes shown from left to right. Regions of loss are indicated by shades of blue, and gains are indicated by shades of red.

See also Figures S1, S2, S3, S4, S5, S6, and S7 and Tables S1A, S1B, S1E, and S2.

rearrangement 3′ of the ETV1 locus with a region on chromosome 14 near the *MIPOL1* gene adjacent to *FOXA1*. This event is similar to previously described *ETV1* translocations in LNCaP and MDA-PCa2b cell lines and in patient samples (Tomlins et al., 2007; Gasi et al., 2011). Overall, while fusions in the four genes were mostly mutually exclusive, three tumors showed evidence for fusions involving more than one of these genes (Table S1E). Given that histologically defined single tumor foci have been shown to be rarely composed of different ETS fusion-positive clones (Cooper et al., 2015; Kunju et al., 2014; Pflueger et al., 2011), it is likely these cases reflect convergent phenotypic evolution in clonally heterogeneous tumors. Tumors defined by *SPOP* mutations were mutually exclusive with all ETS fusion-positive cases, though four of the *SPOP* mutant tumors also possessed *FOXA1* mutations. In all four of these tumors, both the *SPOP* and *FOXA1* mutations were clonal, indicating that they are present in the same tumor cells.

Beyond the class-defining lesions, there were multiple patterns of both known and novel concurrent alterations in key prostate cancer genes. The former included the preponderance of *PTEN* deletions in *ERG* fusion-positive cases (Taylor et al., 2010). Similarly, SPOP mutations have previously been found to occur in ~10% of clinically localized prostate cancers, were mutually exclusive of tumors defined by ETS rearrangements, and may designate a distinct molecular class of disease based

primarily on distinctive SCNA profiles (including deletion of *CHD1*, 6q, and 2q) (Barbieri et al., 2012; Blattner et al., 2014). Beyond reaffirming these known patterns, our taxonomy revealed new relationships and subtypes. Specifically, the *SPOP* mutant/*CHD1*-deleted subset of prostate cancers had notable molecular features, including elevated levels of DNA methylation, homogeneous gene expression patterns, and frequent overexpression of *SPINK1* mRNA, supporting *SPOP* mutation as a key feature in the molecular taxonomy of prostate cancer. Interestingly, mRNA, copy-number, and methylation profiles were similar in tumors with *FOXA1* mutations and those with *SPOP* mutations. Furthermore, we identified a new genomically distinct subtype of prostate cancer defined by hotspot mutations in IDH1, described in greater depth below.

Despite this detailed molecular taxonomy of primary prostate cancers, 26% of all tumors studied appeared to be driven by still-occult molecular abnormalities or by one or more frequent alterations that co-occur with the genomically defined classes. Some of these tumors showed a high burden of copy-number alterations or DNA hypermethylation. Enrichment analysis indicated that this subset of tumors was enriched for mutations in *TP53*, *KDM6A*, and *KMT2D*; deletions of chromosomes 6 and 16; and amplifications of chromosomes 8 (spanning *MYC*) and 11 (*CCND1*) (Table S2). To characterize this group further, we performed whole-genome sequencing of 19 tumor specimens and
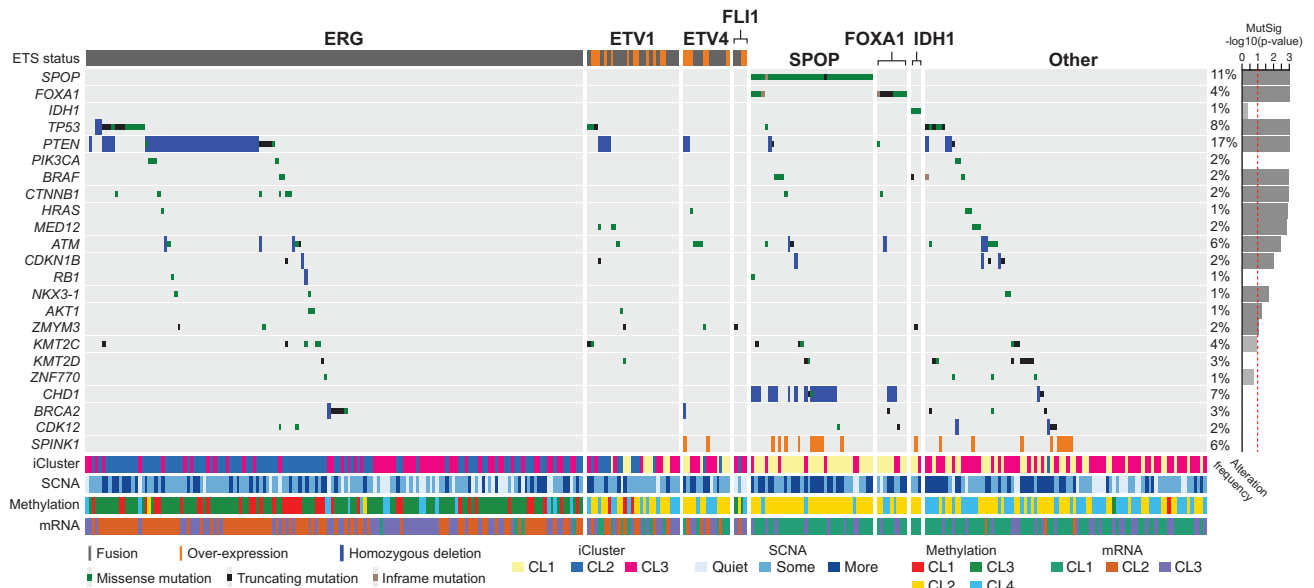
Figure 2. Recurrent Alterations in Primary Prostate Cancer

The spectrum and type of recurrent alterations and genes (mutations, fusions, deletions, and overexpression) in the cohort are shown (left to right) grouped by the molecular subtypes defined in Figure 1. On the right, the statistical significance of individual mutant genes (MutSig q value) is shown. Mutations in *IDH1*, *PIK3CA*, *RB1*, *KMT2D*, *CHD1*, *BRCA2*, and *CDK12* are also shown, despite their not being statistically significant. *SPINK1* overexpression is shown for reference. See also Tables S1B, S1C, S1D, and S1E.

their matched normal tissues, a subset of which had high tumor cellularity but still lacked DNA copy-number alterations or any known or presumed driver lesions. Interestingly, no occult driver abnormalities or highly recurrent regulatory mutations were identified, such as the *TERT* promoter mutation common to many other tumor types (Khurana et al., 2013). Therefore, a significant (up to 26%) subset of primary prostate cancers of both good and poor clinical prognosis (including those with Gleason scores of >8) is driven by as-yet-unexplained molecular alterations.

mRNA clusters were tightly correlated with ETS fusion status, where mRNA cluster 1 consisted primarily of ETS-negative tumors and mRNA clusters 2 and 3 were split among ETS fusion-positive tumors (Figures 1 and S4). miRNA clustering showed a similar pattern, revealing a general difference in miRNA expression between ETS-positive and -negative tumors (Figures 1 and S6). Clustering of RPPA data identified three distinct subgroups, with cluster 3 exhibiting elevated PI3K/AKT, MAP-kinase, and receptor tyrosine kinase activity (Figure S7A). The cluster was not enriched, however, in genomic alterations in these pathways, and in general, there was little correlation of increased pathway activity (as measured by phospho-AKT and other downstream phospho-proteins) with the frequent genomic alterations in the pathways (see the example of *PTEN* deletions in Figure S7B).

Recurrently Altered Genes and Their Patterns across Subtypes

The overall mutational burden of the cohort, inferred from whole-exome sequencing, was 0.94 mutations per megabase (median, range 0.04–28 per megabase), which corresponds to 19 non-synonymous mutations per tumor genome (median; 13–25, 25th and 75th, percentiles respectively). This is consistent

with prior exome and genome-scale sequencing results for localized prostate cancers (Barbieri et al., 2012; Baca et al., 2013) and is lower than the mutational burden of metastatic prostate cancers (Gundem et al., 2015; Grasso et al., 2012; Robinson et al., 2015). These results reaffirm that prostate cancer possesses a lower mutational burden than many other epithelial tumor types that are not associated with a strong exogenous mutagen (Alexandrov et al., 2013; Lawrence et al., 2013). Prior exome sequencing of 112 prostate cancers identified 12 recurrently mutated genes through focused assessment of point mutations and short insertions and deletions (Barbieri et al., 2012). By comparison, mutational significance analysis of these 333 tumor-normal pairs by MutSigCV (Lawrence et al., 2013, 2014) identified 13 significantly mutated genes (q value < 0.1), seven of which had not been previously identified (Figure 2 and Tables S1B and S1C). Among the significantly mutated genes, *SPOP*, *TP53*, *FOXA1*, *PTEN*, *MED12*, and *CDKN1B* were previously identified as recurrently mutated. Additional clinically relevant genes were identified with lower mutation frequencies; these included genes within canonical kinase signaling pathways (*BRAF*, *HRAS*, *AKT1*), the beta-catenin pathway (*CTNNB1*), and the DNA repair pathway (*ATM*). The rate of *BRAF* mutations (2.4%) seen in this study is higher than previously reported; these include several known activating mutations but, curiously, not the canonical V600E hotspot. We identified no *BRAF* fusions, which had previously been reported in a subset of clinically advanced prostate cancer (Palanisamy et al., 2010). *NKX3-1*, previously implicated in familial prostate cancer syndromes and often found to be deleted, was also somatically mutated in this cohort (1% of tumors). While its functional significance is unknown, *ZMYM3*, an epigenetic regulatory protein not previously

implicated in prostate cancer but infrequently mutated in Ewing sarcomas (Tirode et al., 2014) and various pediatric cancers (Huether et al., 2014), was also recurrently mutated (2% of tumors). Genes with known biological relevance that were mutated at frequencies just below the threshold of significance (q value < 0.01) included *KMT2C* (*MLL3*), *KMT2D* (*MLL4*), *APC*, *IDH1*, and *PIK3CA* (Figure 2 and Tables S1B and S1C). Mutations in the tumor suppressor genes *KMT2C*, *KMT2D*, and *APC* were mostly truncating; the *IDH1* and *PIK3CA* mutations occurred in previously characterized hotspots and thus may have therapeutic relevance for those occasional tumors with these mutations.

Notwithstanding these key somatic mutations, the most frequent molecular abnormalities involved chromosomal arm-level copy-number alterations (Taylor et al., 2010). These alterations included recurrent genomic gains of chromosome 7 and 8q and heterozygous losses of 8p, 13q, 16q, and 18 (Figure S3A). Significance analysis of recurrent focal DNA copy-number alterations revealed 20 amplifications and 35 deletions (q value < 0.25, GISTIC 2.0; Figure S3A and Table S1D). Recurrent focal amplifications included those spanning known oncogenes such as *CCND1* (11q13.2, 2%), *MYC* (8q24.21, 8%), and *FGFR1* and *WHSC1L1* (8p11.23, 8%). Recurrent focal deletions were much more common. Homozygous deletions spanning the *PTEN* locus occurred at one of the highest rates of any tumor type studied thus far (15%). Focal deletions of the region between the *TMPRSS2* and *ERG* genes on 21q22.3, which result in TMPRSS2-ERG fusions, were unique to prostate cancers, as expected. Other focal deletions include those spanning tumor suppressors *TP53* (17p13.1), *CDKN1B* (12p13.1), and *MAP3K1* (5q11.2), *FANCD2* (3p26), as well as *SPOPL* (2q22.1) and the complex locus spanning *FOXP1/RYBP/SHQ1* (3p13). *MAP3K7* (6q.12–22) was also frequently deleted, along with deletion of *CHD1* (5q15–q21); co-deletion of these loci has been associated with aggressive ETS-negative prostate cancer (Kluth et al., 2013; Rodrigues et al., 2015).

As the pattern and extent of SCNAs in prostate cancer genomes have been associated with probability of disease recurrence and metastasis in primary prostate cancers (Taylor et al., 2010; Hieronymus et al., 2014; van Dekken et al., 2004; Paris et al., 2004), we sought to identify similar structure in the burden of SCNAs by performing hierarchical clustering of arm-level alterations. We identified three major groups of prostate cancers, one with mostly unaltered genomes (hereafter referred to as *quiet*), a second group encompassing 50% of all tumors with an intermediate level of SCNAs, and a third group with a high burden of arm level genomic gains and losses (Figures S3B and S3C). While a formal outcome analysis was not possible due to the limited clinical follow-up available for this cohort, the subset of tumors with the greatest burden of SCNAs had significantly higher Gleason scores and PSA levels than the other two groups (Figures S3B–S3D). The tumors in this group also had significantly higher tumor cellularity (Figure S3C).

Epigenetic Changes Define Molecularly Distinct Subtypes of Prostate Cancer

Integrative analysis of genetic and epigenetic changes revealed a diversity of DNA methylation changes that defined molecularly distinct subsets of primary prostate cancer (Figure 3). Unsupervised hierarchical clustering of the most variably hypermethylated CpGs identified four epigenetically distinct groups of prostate cancers (Figures S5A and S5B). When integrated with the molecular taxonomy defined above, we found a number of striking associations. Among these was a notable pattern within *ERG* fusion-positive tumors. Specifically, while nearly two-thirds of all *ERG* fusion-positive tumors belonged to an unsupervised cluster with only moderately elevated DNA methylation (DNA methylation cluster 3), the remaining *ERG* fusion-positive tumors comprised a distinct hypermethylated cluster (cluster 1) that was almost exclusively associated with *ERG* fusions. On average, this cluster contained twice the number of hypermethylated loci as DNA methylation cluster 3 (Figure S5A), and the epigenetic patterns were largely distinct from those of *ETV1* and *ETV4* fusion-positive tumors, which showed more heterogeneous methylation. What drives these epigenetically distinct groups of ETS fusion-positive tumors is unknown, but there is considerable diversity in their DNA methylation profiles that may reflect altered epigenetic silencing (Figures S5A and S5B). Together, these results support further ETS fusion-based subtyping of disease but also reveal a greater molecular and likely biological diversity among *ERG* fusion-positive tumors than previously appreciated. Likewise, these results are consistent with in vivo mouse modeling and expression profiling studies that suggest important molecular and clinicopathological differences between *ERG* and non-*ERG* ETS fusion-positive tumors (Baena et al., 2013; Tomlins et al., 2015).

SPOP and *FOXA1* mutant tumors exhibited homogeneous epigenetic profiles. These tumors belonged almost exclusively to DNA methylation cluster 2, a group that also contained a majority of the *ETV1* and *ETV4* but not *ERG*-positive tumors. Lastly, the *IDH1* mutant tumors were notable given their strongly elevated levels of genome-wide DNA hypermethylation (Figure S5B). While of low incidence, these *IDH1* R132 mutant tumors defined a distinct subgroup of what appears to be early-onset prostate cancer (Figure 3B) that possesses fewer DNA copy-number alterations (see Figure 1) or other canonical genomic lesions that are common to most other prostate cancers. *IDH1* and *IDH2* mutations have been associated with a DNA methylation phenotype in other tumor types, most notably in gliomas (Noushmehr et al., 2010) and acute myeloid leukemias (AML, Figueroa et al., 2010). Curiously, *IDH1* mutant prostate cancers possessed even greater levels of genome-wide hypermethylation than either glioma or AML *IDH1* mutant tumors (Figure 3B). After further investigating DNA methylation differences between IDH mutant and wild-type tumors among prostate cancers, gliomas, and AMLs, we found that hypermethylated loci were specific to the cancer type rather than IDH mutants (Figure S5F).

Integrating these epigenetic data with mRNA expression levels, we identified 164 genes that were epigenetically silenced in subsets of the cohort (Figure S5C and Table S1F). These silenced genes were significantly enriched for genes previously found to be differentially expressed in prostate cancer—specifically, genes that are downregulated in metastatic prostate cancer (Chandran et al., 2007) and genes involved in prostate organ development (Schaeffer et al., 2008) (q value < 2.0×10^{-5}). These 164 silenced genes displayed heterogeneous frequencies of
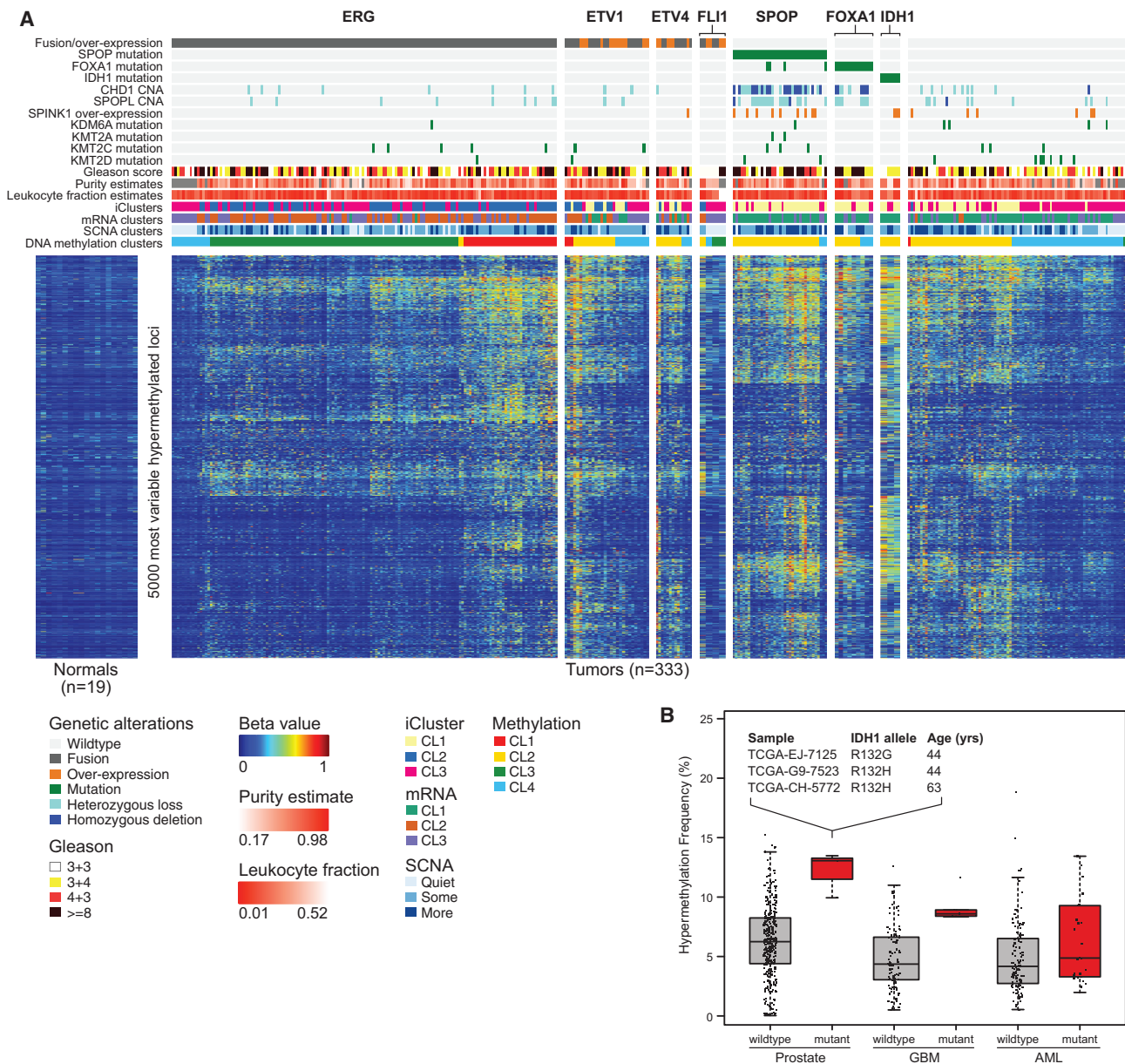
Figure 3. Hypermethylation Is Common across Primary Prostate Cancer

(A) Primary prostate cancers show diverse methylation changes compared to normal prostate samples (left). Unsupervised clustering was performed on the beta-values of the 5,000 most hypermethylated loci, and the results mapped to the genomic subtypes. *ERG*-positive tumors had a high diversity of methylation changes, with a distinct subgroup (cluster 1) nearly unique to this group. *SPOP* and *FOXA1* mutant tumors also exhibited global hypermethylation.

(B) *IDH1* mutant prostate cancers, which are associated with younger age, are among the most hypermethylated tumors, as in glioblastoma (GBM) and AML.

See also Figure S4 and Table S1F.

epigenetic silencing across the cohort. For example, *SHF*, *FAXDC2*, *GSTP1*, *ZNF154*, and *KLF8* were epigenetically silenced in almost all tumors (>85%) whereas *STAT6* was silenced predominantly in ETS fusion-positive tumors and not in *SPOP* and *IDH1* mutant tumors. Conversely, *HEXA* was silenced preferentially in *SPOP* mutant tumors compared to *ERG* fusion-positive tumors (86.5 versus 14.5%, respectively, p < 5.4 × 10^{-15}). Consistent with their increased DNA hyperme-

thylation, the *IDH1* mutant prostate tumors also possessed the greatest number of epigenetically silenced genes among all prostate tumors (Table S1F).

AR Activity Is Variable in Primary Prostate Cancers

The androgen receptor (AR) regulates normal prostate development, as well as critical growth and survival programs in prostate carcinoma. Primary prostate cancer is androgen dependent, and
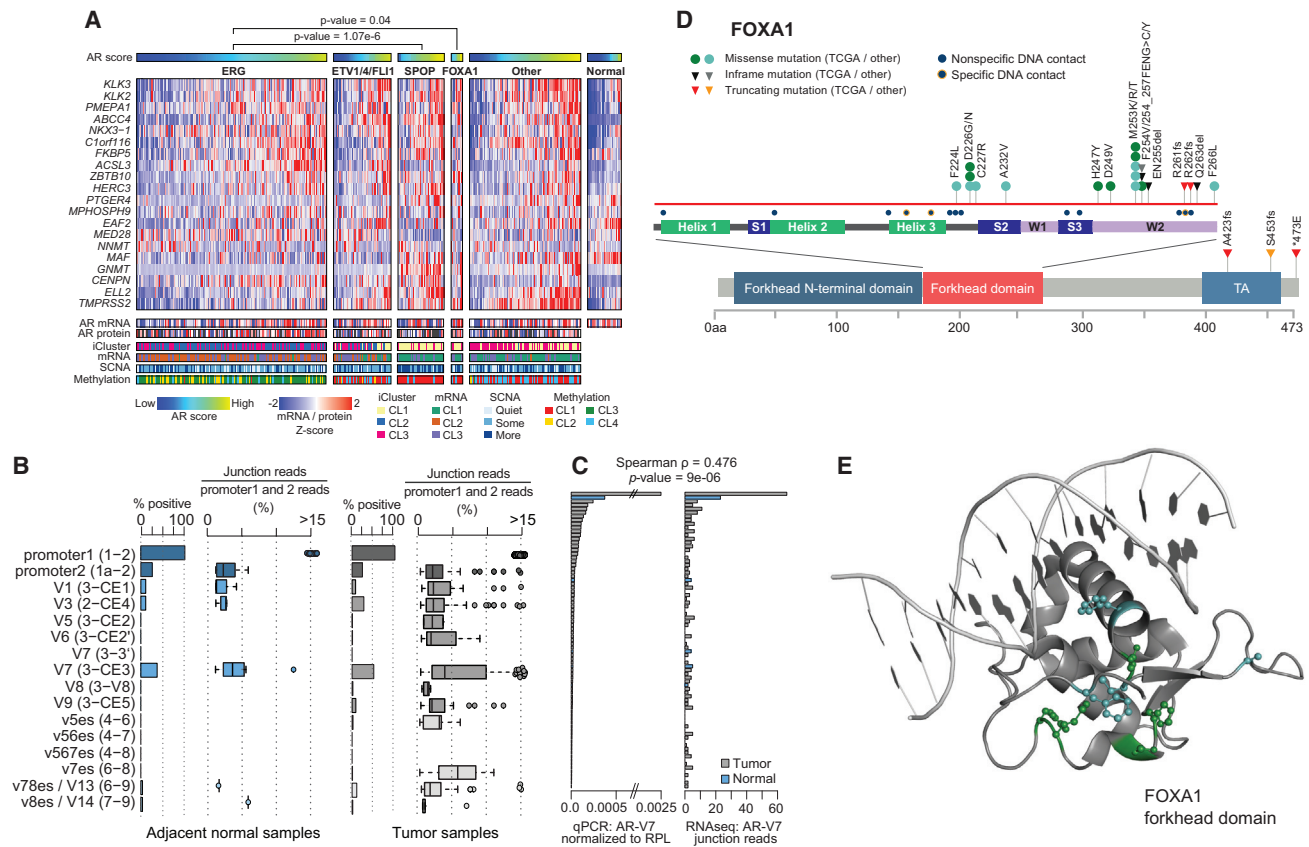
Figure 4. The Diversity of Androgen Receptor Activity in Primary Prostate Cancer

(A) Androgen receptor activity, as inferred by the induction of *AR* target genes, was significantly increased in *SPOP* and *FOXA1* mutant tumors when compared to normal prostate and *ERG*-positive tumors. This increase in activity cannot be fully explained by AR mRNA or protein levels.

(B) Multiple known *AR* splice variants were detected in benign prostate (left) and primary prostate cancer (right), with the AR-V7 variant detected in 50% of tumors.

(C) Real-time qPCR comparison of AR-V7 in 74 tumor samples (gray) and 5 adjacent-normal samples (blue).

(D and E) (D) *FOXA1* missense mutations were clustered in the forkhead domain, mostly in residues that do not form contacts with DNA (see also the 3D structure in panel E).

androgen activity is a central axis in prostate cancer pathogenesis, driving the creation and overexpression of most ETS fusion genes (Lin et al., 2009; Mani et al., 2009; Tomlins et al., 2005). However, the extent to which individual primary prostate cancers differ in androgen sensitivity or dependence is unknown, and the issue has translational implications because AR targeting is therapeutically important. To address these questions, we sought to infer the AR output of tumors by calculating an AR activity score from the expression pattern of 20 genes that are experimentally validated AR transcriptional targets (Hieronymus et al., 2006). This score suggested that a broad spectrum of AR activity exists across all prostate tumors, as well as between genomic subtypes (Figure 4A). Although ETS fusion genes are under AR control, the ETS fusion-positive groups had variable AR transcriptional activity. In contrast, we found that tumors with *SPOP* or *FOXA1* mutations had the highest AR transcriptional activity of all genotypically distinct subsets of prostate cancer (p = 1.1 × 10⁻⁶ and 0.04, respectively, t test). Consistent with this, *SPOP* mutations have been previously implicated in androgen signaling in model systems, since both AR and AR

coactivators are substrates deregulated by *SPOP* mutation (Geng et al., 2013; An et al., 2014; Geng et al., 2014), providing a possible explanation for the associated increase in AR activity seen in this subtype of prostate cancers.

While AR transcriptional output is a proxy for ligand-driven AR activity in many tumors, *AR* transcript variants have been described that encode truncated AR proteins that lack the ligand-binding domain and hence are capable of activating AR target genes in the absence of androgens (Dehm et al., 2008; Watson et al., 2010). Using RNA sequencing reads that spanned the splice junctions unique to each *AR* variant, we quantified the expression of these *AR* transcript variants. This analysis revealed that several *AR* splice variants, most notably AR-V7, can be detected at low levels in primary tumors and, in a few cases, in adjacent benign prostate tissue (Figure 4B), and we validated these expression levels with qPCR (Figure 4C). However, their expression was not associated with differential expression of known AR target genes or with the seven previously defined genomic subtypes. Most detected splice forms were truncated after the DNA-binding domain by the presence
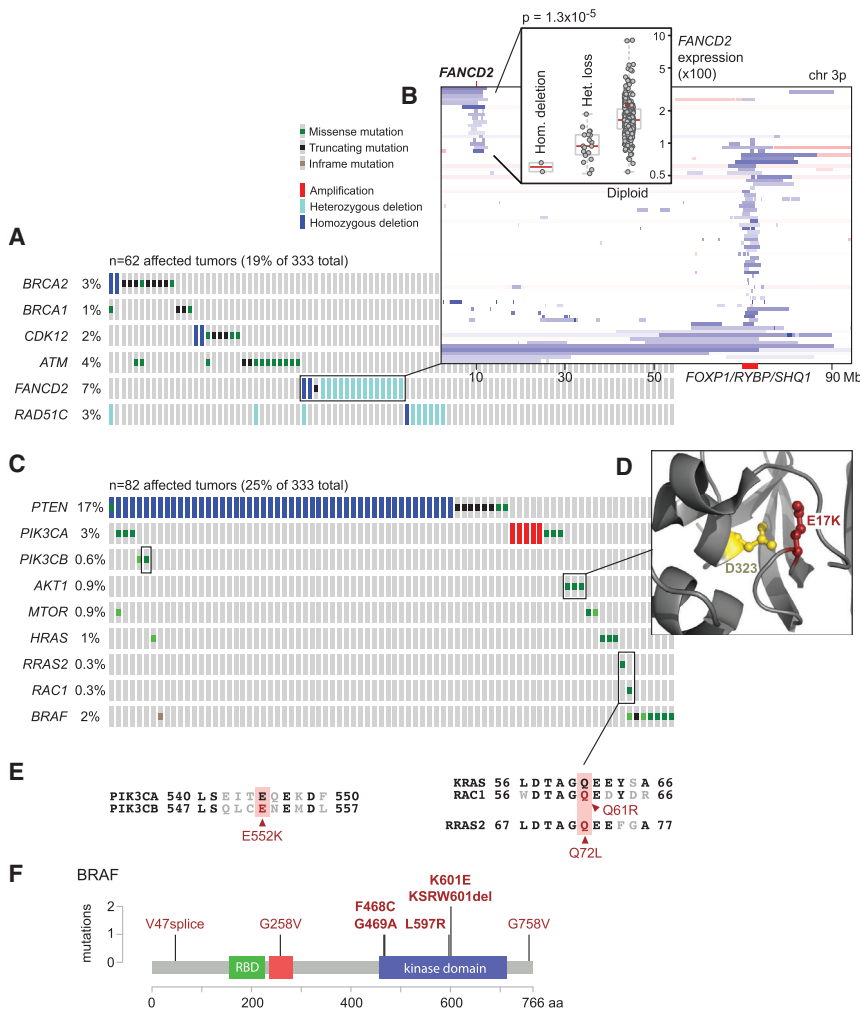
Figure 5. Alterations in Clinically Relevant Pathways

(A) Alterations in DNA repair genes were common in primary prostate cancer, affecting almost 20% of samples through mutations or deletions in *BRCA2*, *BRCA1*, *CDK12*, *ATM*, *FANCD2*, or RAD51C.

(B) Focal deletions of *FANCD2* were found in 7% of samples and were associated with reduced mRNA expression of *FANCD2*.

(C) The RAS or PI-3-Kinase pathways were altered in about a quarter of tumors, mostly through deletion or mutation of *PTEN*, but also through rare mutations in other pathway members.

(D) *AKT1* mutations were found in three samples. Two of them were the known activating E17K, and the third one affected the D323 residue, which is adjacent to E17 in the protein structure.

(E) One of the observed *PIK3CB* mutations, E552K, is paralogous to the known activating E545K mutation in *PIK3CA*, and the *RAC1* Q61 and *RRAS2* Q72 mutations are paralogous to the Q61 mutations in KRAS.

(F) BRAF mutations were found in 2% of samples, mostly in known non-V600E hotspots in the kinase domain.

See also Figure S3.

alterations that define the genomic subclasses described here. While there were some truncating mutations near the C terminus and the C-terminal part of the forkhead domain, the majority of the mutations found here and in other prostate cancer cohorts were missense mutations that primarily affect the winged-helix DNA binding domain of FOXA1. Curiously, these mutations do not directly alter FOXA1 DNA-binding residues (Figures 4D and 4E), a pattern similar to the FOXA1 mutations recently found in lobular breast cancers (TCGA, unpublished data), which suggests that the impact of FOXA1 mutations has less to do with altering DNA binding than with disrupting or altering interactions with other chromatin-bound cofactors.

of a cryptic exon rather than by skipping those exons encoding the ligand-binding domain. Truncated *AR* splice variants were previously assumed to be expressed primarily in metastatic castration-resistant prostate cancers, where, at least for AR-V7, their presence was associated with resistance to hormone therapy (Antonarakis et al., 2014). Hence, our finding that they are expressed in hormone-naive primary prostate cancers is notable.

In prostate cancers, the degree of AR pathway output is controlled not only by AR mRNA and protein expression levels, but also by expression of and mutations in AR cofactors (Heemers and Tindall, 2007). It is therefore notable that *FOXA1* was recurrently mutated in our cohort, as it is a pioneering transcription factor that targets AR and has a demonstrated role in prostate cancer oncogenesis (Jin et al., 2013). We identified *FOXA1* mutations in 4% of the primary prostate cancers studied here, which is similar to the mutation frequency observed previously (Barbieri et al., 2012; Grasso et al., 2012) (Figure 4A). While a subset of these mutations was present in tumors that also possessed *SPOP* mutations and had elevated levels of AR output, *FOXA1* mutations were mutually exclusive with all other

Clinically Actionable DNA Repair Defects in Primary Prostate Cancers

Prior data indicate that several DNA repair pathways are disrupted in a subset of prostate cancers (Karanika et al., 2014; Pritchard et al., 2014). Moreover, the PARP inhibitor olaparib is effective in some patients with prostate cancer (Mateo et al., 2014). Here, we found inactivation of several DNA repair genes that collectively affected 19% of affected individuals (Figure 5A). While we found only one inactivating *BRCA1* germline mutation, a frameshift at V923 caused by a 4 bp deletion (Clinvar RCV000083190.3), *BRCA2* inactivation affected 3% of tumors, including both germline and somatic truncating mutations. All six *BRCA2* germline mutations were K3326*, a C-terminal truncating mutation with debated functional impact but increased prevalence in several tumor types (Farrugia et al., 2008; Martin

et al., 2005; Delahaye-Sourdeix et al., 2015). Two additional tumors possessed focal *BRCA2* homozygous deletions that were accompanied by very low *BRCA2* transcript expression. Four tumors (1%) possessed either loss-of-function mutations or homozygous deletion of *CDK12*, a gene that has been implicated in DNA repair by regulating expression levels of several DNA damage response genes (Blazek et al., 2011) and is recurrently mutated in metastatic prostate cancer (Grasso et al., 2012). *ATM*, an apical kinase of the DNA damage response, which is activated by the Mre11 complex and mediates downstream checkpoint signaling, was affected by a nonsense mutation in one case and by a likely kinase-dead hotspot N2875 mutation in two cases. *FANCD2* was similarly affected by diverse uncommon lesions, including a truncating mutation in one tumor, homozygous deletion in two tumors, and focal heterozygous losses in 6% of the cohort (Figure 5B). *RAD51C* (3%) was affected by focal DNA losses, most of which were heterozygous. Finally, it was notable that heterozygous losses of *BRCA2* (13q13.1) almost always coincided with concurrent loss of the distant *RB1* tumor suppressor gene (13q14.2) (Figure S3D). The observation that nearly 20% of primary prostate cancers bear genomic defects involving DNA repair pathways is remarkably consistent with the recently announced TOPARP-A Phase II trial results in patients with metastatic castration-resistant prostate cancer, indicating that clinical responses to the PARP inhibitor olaparib likely occurred in the subgroup of tumors bearing defects in DNA repair genes (Mateo et al., 2014; Robinson et al., 2015).

Clinically Actionable Lesions in PI3K and Ras Signaling

The long tail of the frequency distribution of molecular abnormalities is particularly notable among primary prostate cancers. Beyond *PTEN*, which was deleted or mutated in 17% of the cohort, various driver mutations in effectors of PI3K signaling were present at low incidence (Figure 5C). *PIK3CA*, which encodes the 110 kDa catalytic subunit of phosphatidylinositol 3-kinase, was mutated in six tumors, including one case possessing coincident activating mutations (E542A and N345I), both of which appeared to be subclonal. The other four *PIK3CA* mutations were all known activating mutational hotspots (E545K, Q546K, N345I, and C420R), while one had a mutation of unknown function (E474A). Focal *PIK3CA* amplification with associated mRNA overexpression occurred in ~1% of cases. Interestingly, *PIK3CB* was mutated in two tumors that also possessed coincident homozygous deletions of *PTEN*, both of which were clonal. *PIK3CB* E552K was found in one tumor at a paralogous residue to the canonical *PIK3CA* helical domain E545K mutant and is presumably activating (Figure 5E). As *PTEN*-deleted tumors are likely *PIK3CB*-dependent due to the feedback inhibition of *PIK3CA*, co-existent loss and mutation of *PTEN* and *PIK3CB* may be elevating PI3K pathway output and perhaps indicating a set of tumors in which combined PI3K and androgen signaling inhibition may be effective (Schwartz et al., 2015). Among other lesions that drive PI3K signaling, *AKT1* was mutated in three tumors. Two tumors had the known E17K hotspot mutation, while another encoded a D323Y mutation. Whereas E17K is the most common hotspot in *AKT1* across human cancer, the D323Y variant is uncommon, having been identified previously in one lung adenocarcinoma (Cancer Genome Atlas Research Network, 2014) and

one urothelial bladder cancer (Guo et al., 2013). Nevertheless, while distant linearly from the activating E17K hotspot, in three dimensions, this D323Y kinase domain mutant directly abuts the PH-domain containing E17K (Figure 5D) and has been described as potentially activating (Parikh et al., 2012).

We also identified known or presumed driver mutations in several other genes of the MAPK pathway, affecting 25% of the tumors (Figure 5A). *HRAS* was mutated in four tumors, of which three were Q61R hotspot mutations. Two mutations arose in other Ras family small GTPases. While both *RAC1* Q61R and *RRAS2* Q72L occurred only once each, they affected residues paralogous to the RAS Q61 hotspot (Figure 5E) (Chang et al., 2015). We also identified eight *BRAF* mutations, though, curiously, none were the common V600E mutation that is prevalent in cutaneous melanomas, thyroid cancers, and many other tumor types. Five *BRAF* mutations are likely activating, including known hotspots (K601E, G469A, L597R), two of which confer sensitivity to MEK inhibitors (Dahlman et al., 2012; Bowyer et al., 2014). Another mutation was a likely activating in-frame 3 amino acid deletion at K601 (Figure 5F), while the final mutation (F468C) affected the adjacent residue to the known G469 hotspot. Together, these findings reveal a long tail of low-incidence potentially actionable predicted driver mutations present across the molecular taxonomy of prostate cancer.

Comparison with Metastatic Prostate Cancer

To put these results in context, we compared our findings with those from a recently published cohort of 150 castration-resistant metastatic prostate cancer samples (Robinson et al., 2015). The analysis revealed some similarities and many differences between primary and treated metastatic disease. Although the overall burden of copy-number alterations and mutations was significantly higher in the metastatic samples (Figure 6A), consistent with previous findings (Taylor et al., 2010; Grasso et al., 2012), the primary and metastatic samples were remarkably similar in their subtype distribution, with the exception that the metastatic dataset contained no *IDH1* mutant tumors (Figure 6B). We compared the frequencies of all recurrently altered genes described in both studies and found that, similar to the overall burden of genomic alterations (Figure 6A), many genes and pathways have increased alteration rates in the metastatic samples (Figure 6C and Table S3). Androgen receptor signaling was more frequently altered in the metastatic samples, most often by amplification or mutation of *AR*, events that were essentially absent in primary samples. Interestingly, *SPOP* mutations were somewhat less frequent in the metastatic samples (8% versus 11% in the primary samples). DNA repair and PI3K pathway alterations were more frequent in the metastatic samples, as were mutations or deletions of *TP53*, *RB1*, *KMT2C*, and *KMT2D*. Interestingly, we found no focal, clonal *MYCL* amplifications, which were recently described in primary prostate cancer (Boutros et al., 2015), in either dataset nor in a separate set of 63 untreated prostate cancer samples (Hovelson et al., 2015).

DISCUSSION

The comprehensive molecular analyses of primary prostate cancers presented here reveal highly diverse genomic, epigenomic,
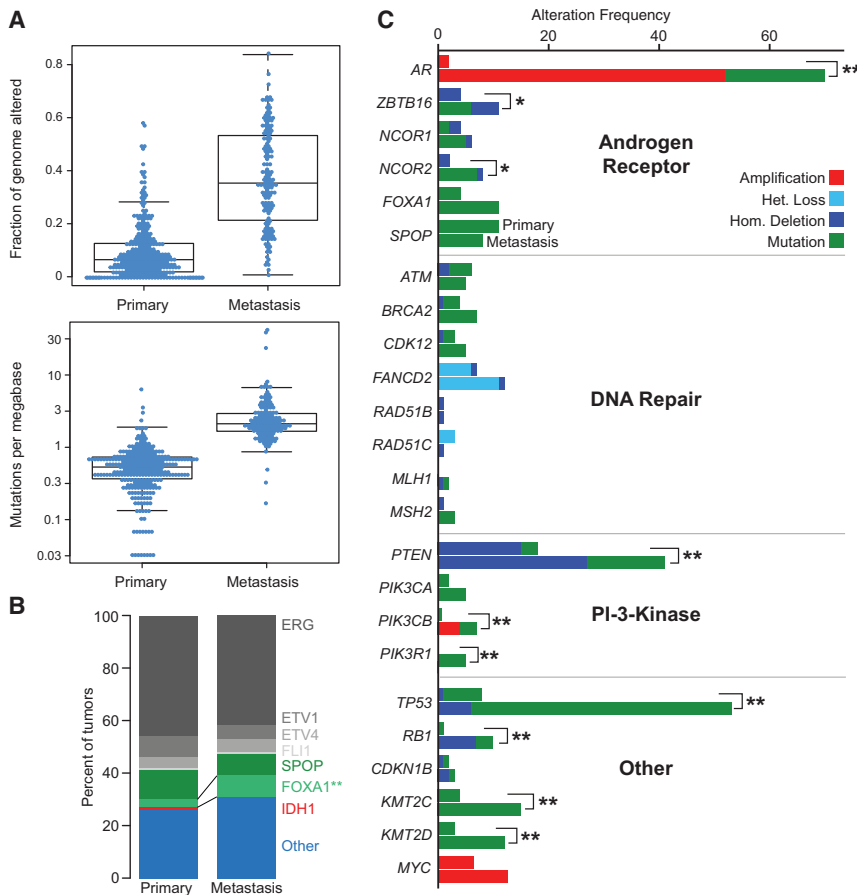
Figure 6. Comparison of Primary with Metastatic Prostate Cancer

(A) Metastatic prostate cancer samples have more copy-number alterations (top, measured as fraction of genome altered) and mutations (bottom).

(B) The relative distribution of main subtypes (*ERG*, *ETV1/4*, *FLI1*, *SPOP*, *FOXA1*, *IDH1*, other) is similar in primary and metastatic samples.

(C) The alteration frequencies of several genes and pathways are higher in metastatic samples. The upper bar for each gene indicates the alteration frequency in primary samples, the lower bar for metastatic samples. The most notable differences in alteration frequencies involve the Androgen Receptor pathway, the PI3K pathway, and *TP53*. See also Table S3.

are all ETS fusion negative and *SPOP* wild-type, have little SCNA burden, and possess elevated levels of genome-wide methylation. The levels of methylation observed in this methylator phenotype are higher than those observed in *IDH1* mutant GBMs and AMLs. Consistent with our observations, a recently published clinical study of 117 prostate cancers identified a single *IDH1* mutant prostate cancer from 56-year-old affected individual that also lacked significant copy number alterations, ETS gene fusions, or driver mutations (Hovelson et al., 2015). Future studies in cohorts with sufficient clinical follow-up will be able to ask whether the *IDH1* mutant prostate cancers are prognostically distinct, as noted for gliomas (Noushmehr et al., 2010) and AMLs (Mardis et al., 2009), and if they are sensitive to newly developed IDH1-targeted therapeutics (Rohle et al., 2013).

Interestingly, 26% of the tumors in this study could not be characterized by one of the taxonomy-defining cardinal genomic alterations. The 26% were clinically and genomically heterogeneous, with some tumors exhibiting extensive DNA copy-number alterations and high Gleason scores indicative of poorer prognosis. About a third of them were genomically similar to *SPOP* and *FOXA1* mutant tumors but lacked any canonical mutation (iCluster 1, methylation cluster 2, mRNA cluster 1); others were enriched for mutations of *TP53*, *KDM6A*, and *KMT2D* or specific SCNAs spanning *MYC* and *CCND1*. Many of the tumors had low Gleason score with few if any DNA copy-number alterations and a normal-like DNA methylation pattern. As previously reported, tumors with fewer genomic alterations were also more commonly Gleason score 6 tumors (38% in the "quiet" class versus 8% in the class with the greatest burden of alterations).

Tumor cellularity, as assessed by pathology review, was lower among tumors with fewer SCNAs (one-sided Mann-Whitney test, p = 0.0002), indicating that the apparent lower burden of alterations in tumors with smaller volumes may be due in part to their tumor purities being lower. However, the lower cellularity of these tumors did not limit the detection of clonal molecular

and transcriptomic patterns. Major subtypes could be defined by fusions of the ETS family genes *ERG*, *ETV*, *ETV4*, or *FLI1* and by mutations in *SPOP*, *FOXA1*, or *IDH1*. However, even within the groups, there was significant diversity in DNA copy-number alterations, gene expression, and DNA methylation. The mutational heterogeneity mirrors the heterogeneous natural history of primary prostate cancers.

Although the broad spectrum of copy-number alterations in tumors with ETS fusions has been previously characterized (Demichelis et al., 2009; Taylor et al., 2010), here we uncovered additional differences between the epigenetic profiles of those tumors. We found that *ERG* fusion-positive tumors can be subdivided into two methylation subtypes: one with lower levels of methylation, and one with a distinct spectrum of hypermethylation. Many genes were epigenetically silenced as a result of the hypermethylation in the latter tumors. While further studies will be required to determine which silencing events are linked to prostate cancer pathogenesis, the findings presented here reveal variability among what was previously considered to be genetically homogeneous prostate cancer subtypes.

We have also identified a distinct subgroup of tumors with *IDH1* R132 mutations that is associated with younger age at diagnosis. Although *IDH1* mutations have previously been identified in prostate cancer with a similar incidence (2.7%) (Ghiam et al., 2012; Kang et al., 2009), we show here that those tumors

lesions since tumor cellularity between ETS fusion-positive and these fusion-negative tumors was not significantly different (two-sided Mann-Whitney test, p = 0.32). One must also keep in mind that this study was limited to a single tumor focus for each affected individual, even though the vast majority of primary prostate tumors are multifocal and molecular heterogeneity between different foci has been demonstrated (Cooper et al., 2015; Boutros et al., 2015; Lindberg et al., 2015). Such issues must be considered when designing new therapeutic approaches and biomarker panels for clinical use, as affected individuals likely have more than one of these molecular subtypes present due to this commonly occurring tumor multifocality and molecular heterogeneity.

Primary prostate cancers exhibit a wide range of androgen receptor activity. This study demonstrates for the first time a direct association between mutations in *SPOP* or *FOXA1* and increased AR-driven transcription in human prostate cancers. Further studies in preclinical models, as well as in clinical trial settings, will be required to understand the implications of variable AR activity in the contexts of chemoprevention and prostate cancer-directed treatment strategies (Mostaghel et al., 2010). Other, more immediately actionable opportunities for targeted therapy exist for the 19% of primary prostate cancers that have defects in DNA repair and for the nearly equal number of cancers with altered key effectors of both PI3K and MAPK pathways. While the numbers of DNA repair defects found in organ-confined prostate tumors may be lower than those found in metastatic prostate cancer (Robinson et al., 2015), an increase in the number of such defects with disease progression suggests a possible advantage to targeting DNA repair-deficient tumors at an earlier stage of disease, perhaps at initial diagnosis. Such strategies may include preventing DNA damage, as well as targeting deficient DNA repair (Ferguson et al., 2015). Alterations in the PI3K/MTOR pathway also play an important role: beyond the frequent inactivation of *PTEN*, we document rare activation of *PIK3CA*, *PIK3CB*, *AKT1*, and *MTOR*, and of several small GTPases, including *HRAS*, as well as *BRAF*. As DNA sequencing of tumor samples becomes more widely adopted earlier in the clinical care of cancer patients, such alterations may emerge as candidates for inclusion in clinical trials after front-line therapy.

In summary, our integrative assessment of 333 primary prostate cancers has confirmed previously defined molecular subtypes across multiple genomic platforms and identified novel alterations and subtype diversity. It provides a resource for continued investigation into the molecular and biological heterogeneity of the most common cancer in American men.

EXPERIMENTAL PROCEDURES

Tumor and matched normal specimens were obtained from prostate cancer patients who provided informed consent and were approved for collection and distribution by local Institutional Review boards. Blocks frozen in OCT were made of all tumors and of paired benign tissue when present. A 5 micron section was cut from both the top and bottom of the OCT block of 111 tumor cases and from the top or bottom only of the OCT block of 222 tumor cases. Out of 39 normal samples included in the freeze, 23 underwent pathology review, and prostate origin (i.e., no seminal vesicles) and absence of tumor and high grade prostate intraepithelial neoplasia (HGPIN) were confirmed. Tissue

images were reviewed by eight genitourinary pathologists, who reported the primary and secondary Gleason patterns of cancer for each slide and estimates of tumor cellularity in 10% increments (from 0%–100%). In case of discrepancies of Gleason scores between the top and bottom sections, the Gleason scores of cancer in the section with the largest area of tumor were used. A subset of 54 cases was reviewed by two pathologists. Discrepancies that occurred between the two pathologists were reconciled by blind review by a third pathologist.

DNA, RNA, and protein were purified and distributed throughout the TCGA network. Samples with evidence for RNA degradation were excluded from the study (Supplemental Experimental Procedures). In total, 333 primary tumors with associated clinicopathologic data were assayed on at least four molecular profiling platforms. Platforms included exome and whole genome DNA sequencing, RNA sequencing, miRNA sequencing, SNP arrays, DNA methylation arrays, and reverse phase protein arrays. Integrated multiplatform analyses were performed.

The data and analysis results can be explored through the Broad Institute FireBrowse portal (http://firebrowse.org/?cohort=PRAD), the cBioPortal for Cancer Genomics (http://www.cbioportal.org/study.do?cancer_study_id=prad_tcga_pub), TCGA Batch Effects (http://bioinformatics.mdanderson.org/tcgambatch/), Regulome Explorer (http://explorer.cancerregulome.org/), and Next-Generation Clustered Heat Maps (http://bioinformatics.mdanderson.org/TCGA/NGCHMPortal/). See also Supplemental Information and the TCGA publication page (https://tcga-data.nci.nih.gov/docs/publications/prad_2015/).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.025.

CONSORTIA

The members of The Cancer Genome Atlas Research Network are Adam Abeshouse, Jaeil Ahn, Rehan Akbani, Adrian Ally, Samirkumar Amin, Christopher D. Andry, Matti Annala, Armen Aprikian, Joshua Armenia, Arshi Arora, J. Todd Auman, Miruna Balasundaram, Saianand Balu, Christopher E. Barbieri, Thomas Bauer, Christopher C. Benz, Alain Bergeron, Rameen Beroukhim, Mario Berrios, Adrian Bivol, Tom Bodenheimer, Lori Boice, Moiz S. Bootwalla, Rodolfo Borges dos Reis, Paul C. Boutros, Jay Bowen, Reanne Bowlby, Jeffrey Boyd, Robert K. Bradley, Anne Breggia, Fadi Brimo, Christopher A. Bristow, Denise Brooks, Bradley M. Broom, Alan H. Bryce, Glenn Bubley, Eric Burks, Yaron S.N. Butterfield, Michael Button, David Canes, Carlos G. Carlotti, Rebecca Carlsen, Michel Carmel, Peter R. Carroll, Scott L. Carter, Richard Cartun, Brett S. Carver, June M. Chan, Matthew T. Chang, Yu Chen, Andrew D. Cherniack, Simone Chevalier, Lynda Chin, Juok Cho, Andy Chu, Eric Chuah, Sudha Chudamani, Kristian Cibulskis, Giovanni Ciriello, Amanda Clarke, Matthew R. Cooperberg, Niall M. Corcoran, Anthony J. Costello, Janet Cowan, Daniel Crain, Erin Curley, Kerstin David, John A. Demchok, Francesca Demichelis, Noreen Dhalla, Rajiv Dhir, Alexandre Doueik, Bettina Drake, Heidi Dvinge, Natalya Dyakova, Ina Felau, Martin L. Ferguson, Scott Frazer, Stephen Freedland, Yao Fu, Stacey B. Gabriel, Jianjiong Gao, Johanna Gardner, Julie M. Gastier-Foster, Nils Gehlenborg, Mark Gerken, Mark B. Gerstein, Gad Getz, Andrew K. Godwin, Anuradha Gopalan, Markus Graefen, Kiley Graim, Thomas Gribbin, Ranabir Guin, Manaswi Gupta, Angela Hadjipanayis, Syed Haider, Lucie Hamel, D. Neil Hayes, David I. Heiman, Julian Hess, Katherine A. Hoadley, Andrea H. Holbrook, Robert A. Holt, Antonia Holway, Christopher M. Hovens, Alan P. Hoyle, Mei Huang, Carolyn M. Hutter, Michael Ittmann, Lisa Iype, Stuart R. Jefferys, Corbin D. Jones, Steven J.M. Jones, Hartmut Juhl, Andre Kahles, Christopher J. Kane, Katayoon Kasaian, Michael Kerger, Ekta Khurana, Jaegil Kim, Robert J. Klein, Raju Kucherlapati, Louis Lacombe, Marc Ladanyi, Phillip H. Lai, Peter W. Laird, Eric S. Lander, Mathieu Latour, Michael S. Lawrence, Kevin Lau, Tucker LeBien, Darlene Lee, Semin Lee, Kjong-Van Lehmann, Kristen M. Leraas, Ignaty Leshchiner, Robert Leung, John A. Libertino, Tara M. Lichtenberg, Pei Lin, W. Marston Linehan, Shiyun Ling, Scott M. Lippman, Jia Liu, Wenbin Liu, Lucas Lochovsky, Massimo

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

REFERENCES

Ahn, J., Yuan, Y., Parmigiani, G., Suraokar, M.B., Diao, L., Wistuba, I.I., and Wang, W. (2013). DeMix: deconvolution for mixed cancer transcriptomes using raw measured data. Bioinformatics 29, 1865–1871.

Al Olama, A.A., Kote-Jarai, Z., Berndt, S.I., Conti, D.V., Schumacher, F., Han, Y., Benlloch, S., Hazelett, D.J., Wang, Z., Saunders, E., et al.; Breast and Prostate Cancer Cohort Consortium (BPC3); PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium; COGS (Collaborative Oncological Gene-environment Study) Consortium; GAME-ON/ELLIPSE Consortium (2014). A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. Nat. Genet. 46, 1103–1109.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Bian-kin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. Nature 500, 415–421.

An, J., Wang, C., Deng, Y., Yu, L., and Huang, H. (2014). Destruction of full-length androgen receptor by wild-type SPOP, but not prostate-cancer-associ-ated mutants. Cell Rep. 6, 657–669.

Antonarakis, E.S., Lu, C., Wang, H., Luber, B., Nakazawa, M., Roeser, J.C., Chen, Y., Mohammad, T.A., Chen, Y., Fedor, H.L., et al. (2014). AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. N. Engl. J. Med. 371, 1028–1038.

Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., Ghandi, M., et al. (2013). Punc-tuated evolution of prostate cancer genomes. Cell 153, 666–677.

Baena, E., Shao, Z., Linn, D.E., Glass, K., Hamblen, M.J., Fujiwara, Y., Kim, J., Nguyen, M., Zhang, X., Godinho, F.J., et al. (2013). ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and pa-tients. Genes Dev. 27, 683–698.

Barbieri, C.E., Baca, S.C., Lawrence, M.S., Demichelis, F., Blattner, M., Theur-illat, J.P., White, T.A., Stojanov, P., Van Allen, E., Stransky, N., et al. (2012). Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat. Genet. 44, 685–689.

Berger, M.F., Lawrence, M.S., Demichelis, F., Drier, Y., Cibulskis, K., Siva-chenko, A.Y., Sboner, A., Esgueva, R., Pflueger, D., Sougnez, C., et al. (2011). The genomic complexity of primary human prostate cancer. Nature 470, 214–220.
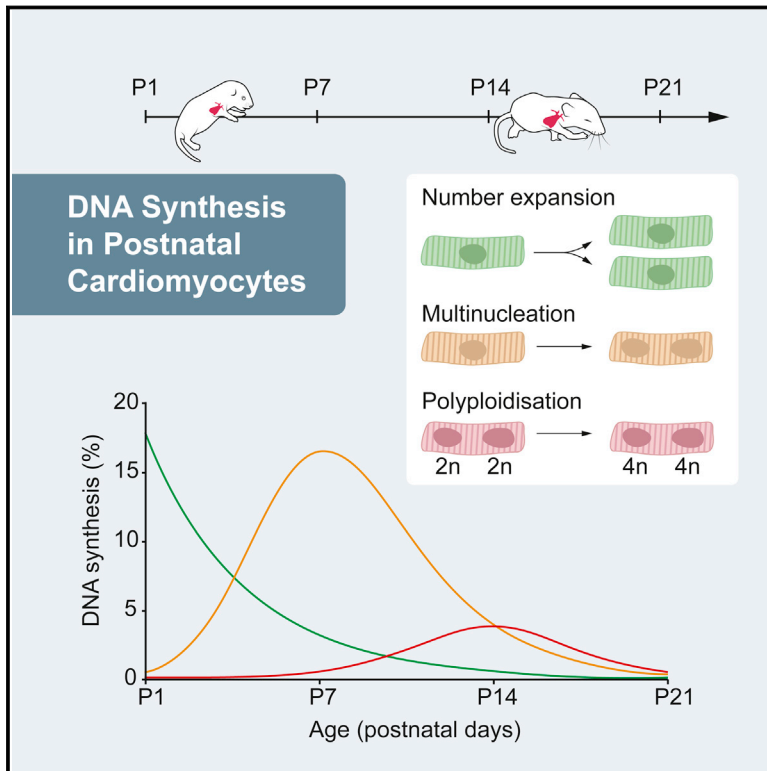
Blattner, M., Lee, D.J., O'Reilly, C., Park, K., MacDonald, T.Y., Khani, F., Turner, K.R., Chiu, Y.L., Wild, P.J., Dolgalev, I., et al. (2014). SPOP mutations in prostate cancer across demographically diverse patient cohorts. Neoplasia 16, 14–20.

Blazek, D., Kohoutek, J., Bartholomeeusen, K., Johansen, E., Hulinkova, P., Luo, Z., Cimermancic, P., Ule, J., and Peterlin, B.M. (2011). The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. Genes Dev. 25, 2158–2172.

Börno, S.T., Fischer, A., Kerick, M., Fälth, M., Laible, M., Brase, J.C., Kuner, R., Dahl, A., Grimm, C., Sayanjali, B., et al. (2012). Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. Cancer Discov. 2, 1024–1035.

Boutros, P.C., Fraser, M., Harding, N.J., de Borja, R., Trudel, D., Lalonde, E., Meng, A., Hennings-Yeomans, P.H., McPherson, A., Sabelnykova, V.Y., et al. (2015). Spatial genomic heterogeneity within localized, multifocal prostate cancer. Nat. Genet. 47, 736–745.

Bowyer, S.E., Rao, A.D., Lyle, M., Sandhu, S., Long, G.V., McArthur, G.A., Raleigh, J.M., Hicks, R.J., and Millward, M. (2014). Activity of trametinib in K601E and L597Q BRAF mutation-positive metastatic melanoma. Melanoma Res. 24, 504–508.

Cancer Genome Atlas Research Network (2014). Comprehensive molecular profiling of lung adenocarcinoma. Nature 511, 543–550.

Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. 30, 413–421.

Chandran, U.R., Ma, C., Dhir, R., Bisceglia, M., Lyons-Weiler, M., Liang, W., Michalopoulos, G., Becich, M., and Monzon, F.A. (2007). Gene expression profiles of prostate cancer reveal involvement of multiple molecular pathways in the metastatic process. BMC Cancer 7, 64.

Chang, M.T., Asthana, S., Gao, S.P., Lee, B.H., Chapman, J.S., Kandoth, C., Gao, J., Socci, N.D., Solit, D.B., Olshen, A.B., et al. (2015). Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat. Biotechnol. Published online November 9, 2015. http://dx.doi.org/10.1038/nbt.3391.

Cooper, C.S., Eeles, R., Wedge, D.C., Van Loo, P., Gundem, G., Alexandrov, L.B., Kremeyer, B., Butler, A., Lynch, A.G., Camacho, N., et al.; ICGC Prostate Group (2015). Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. Nat. Genet. 47, 367–372.

Cooperberg, M.R., Broering, J.M., and Carroll, P.R. (2009). Risk assessment for prostate cancer metastasis and mortality at the time of diagnosis. J. Natl. Cancer Inst. 101, 878–887.

D'Amico, A.V., Whittington, R., Malkowicz, S.B., Schultz, D., Blank, K., Broderick, G.A., Tomaszewski, J.E., Renshaw, A.A., Kaplan, I., Beard, C.J., and Wein, A. (1998). Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. JAMA 280, 969–974.

Dahlman, K.B., Xia, J., Hutchinson, K., Ng, C., Hucks, D., Jia, P., Atefi, M., Su, Z., Branch, S., Lyle, P.L., et al. (2012). BRAF(L597) mutations in melanoma are associated with sensitivity to MEK inhibitors. Cancer Discov. 2, 791–797.

Dehm, S.M., Schmidt, L.J., Heemers, H.V., Vessella, R.L., and Tindall, D.J. (2008). Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. Cancer Res. 68, 5469–5477.

Delahaye-Sourdeix, M., Anantharaman, D., Timofeeva, M.N., Gaborieau, V., Chabrier, A., Vallée, M.P., Lagiou, P., Holcátová, I., Richiardi, L., Kjaerheim, K., et al. (2015). A rare truncating BRCA2 variant and genetic susceptibility to upper aerodigestive tract cancer. J. Natl. Cancer Inst. 107, djv037.

Demichelis, F., Setlur, S.R., Beroukhim, R., Perner, S., Korbel, J.O., Lafargue, C.J., Pflueger, D., Pina, C., Hofer, M.D., Sboner, A., et al. (2009). Distinct genomic aberrations associated with ERG rearranged prostate cancer. Genes Chromosomes Cancer 48, 366–380.

Farrugia, D.J., Agarwal, M.K., Pankratz, V.S., Deffenbaugh, A.M., Pruss, D., Frye, C., Wadum, L., Johnson, K., Mentlick, J., Tavtigian, S.V., et al. (2008). Functional assays for classification of BRCA2 variants of uncertain significance. Cancer Res. 68, 3523–3531.

Ferguson, L.R., Chen, H., Collins, A.R., Connell, M., Damia, G., Dasgupta, S., Malhotra, M., Meeker, A.K., Amedei, A., Amin, A., et al. (2015). Genomic instability in human cancer: Molecular insights and opportunities for therapeutic attack and prevention through diet and nutrition. Semin. Cancer Biol. Published online April 11, 2015. http://dx.doi.org/10.1016/j.semcancer.2015.03.005.

Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2013). GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 (Lyon, France: International Agency for Research on Cancer).

Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell 18, 553–567.

Friedlander, T.W., Roy, R., Tomlins, S.A., Ngo, V.T., Kobayashi, Y., Azameera, A., Rubin, M.A., Pienta, K.J., Chinnaiyan, A., Ittmann, M.M., et al. (2012). Common structural and epigenetic changes in the genome of castration-resistant prostate cancer. Cancer Res. 72, 616–625.

Gasi, D., van der Korput, H.A., Douben, H.C., de Klein, A., de Ridder, C.M., van Weerden, W.M., and Trapman, J. (2011). Overexpression of full-length ETV1 transcripts in clinical prostate cancer due to gene translocation. PLoS ONE 6, e16332.

Geng, C., He, B., Xu, L., Barbieri, C.E., Eedunuri, V.K., Chew, S.A., Zimmermann, M., Bond, R., Shou, J., Li, C., et al. (2013). Prostate cancer-associated mutations in speckle-type POZ protein (SPOP) regulate steroid receptor coactivator 3 protein turnover. Proc. Natl. Acad. Sci. USA 110, 6997–7002.

Geng, C., Rajapakshe, K., Shah, S.S., Shou, J., Eedunuri, V.K., Foley, C., Fiskus, W., Rajendran, M., Chew, S.A., Zimmermann, M., et al. (2014). Androgen receptor is the key transcriptional mediator of the tumor suppressor SPOP in prostate cancer. Cancer Res. 74, 5631–5643.

Ghiam, A.F., Cairns, R.A., Thoms, J., Dal Pra, A., Ahmed, O., Meng, A., Mak, T.W., and Bristow, R.G. (2012). IDH mutation status in prostate cancer. Oncogene 31, 3826.

Gopalan, A., Leversha, M.A., Satagopan, J.M., Zhou, Q., Al-Ahmadie, H.A., Fine, S.W., Eastham, J.A., Scardino, P.T., Scher, H.I., Tickoo, S.K., et al. (2009). TMPRSS2-ERG gene fusion is not associated with outcome in patients treated by prostatectomy. Cancer Res. 69, 1400–1406.

Grasso, C.S., Wu, Y.M., Robinson, D.R., Cao, X., Dhanasekaran, S.M., Khan, A.P., Quist, M.J., Jing, X., Lonigro, R.J., Brenner, J.C., et al. (2012). The mutational landscape of lethal castration-resistant prostate cancer. Nature 487, 239–243.

Gundem, G., Van Loo, P., Kremeyer, B., Alexandrov, L.B., Tubio, J.M., Papaemmanuil, E., Brewer, D.S., Kallio, H.M., Högnäs, G., Annala, M., et al.; ICGC Prostate UK Group (2015). The evolutionary history of lethal metastatic prostate cancer. Nature 520, 353–357.

Guo, G., Sun, X., Chen, C., Wu, S., Huang, P., Li, Z., Dean, M., Huang, Y., Jia, W., Zhou, Q., et al. (2013). Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. Nat. Genet. 45, 1459–1463.

Heemers, H.V., and Tindall, D.J. (2007). Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. Endocr. Rev. 28, 778–808.

Hieronymus, H., Lamb, J., Ross, K.N., Peng, X.P., Clement, C., Rodina, A., Nieto, M., Du, J., Stegmaier, K., Raj, S.M., et al. (2006). Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. Cancer Cell 10, 321–330.

Hieronymus, H., Schultz, N., Gopalan, A., Carver, B.S., Chang, M.T., Xiao, Y., Heguy, A., Huberman, K., Bernstein, M., Assel, M., et al. (2014). Copy number alteration burden predicts prostate cancer relapse. Proc. Natl. Acad. Sci. USA *111*, 11139–11144.

Hovelson, D.H., McDaniel, A.S., Cani, A.K., Johnson, B., Rhodes, K., Williams, P.D., Bandla, S., Bien, G., Choppa, P., Hyland, F., et al. (2015). Development and validation of a scalable next-generation sequencing system for assessing relevant somatic variants in solid tumors. Neoplasia *17*, 385–399.

Huether, R., Dong, L., Chen, X., Wu, G., Parker, M., Wei, L., Ma, J., Edmonson, M.N., Hedlund, E.K., Rusch, M.C., et al. (2014). The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. Nat. Commun. *5*, 3630.

Jin, H.J., Zhao, J.C., Ogden, I., Bergan, R.C., and Yu, J. (2013). Androgen receptor-independent function of FoxA1 in prostate cancer metastasis. Cancer Res. *73*, 3725–3736.

Kang, M.R., Kim, M.S., Oh, J.E., Kim, Y.R., Song, S.Y., Seo, S.I., Lee, J.Y., Yoo, N.J., and Lee, S.H. (2009). Mutational analysis of IDH1 codon 132 in glioblastomas and other common cancers. Int. J. Cancer *125*, 353–355.

Karanika, S., Karantanos, T., Li, L., Corn, P.G., and Thompson, T.C. (2014). DNA damage response and prostate cancer: defects, regulation and therapeutic implications. Oncogene *34*, 2815–2822.

Kattan, M.W., Eastham, J.A., Stapleton, A.M., Wheeler, T.M., and Scardino, P.T. (1998). A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. J. Natl. Cancer Inst. *90*, 766–771.

Khurana, E., Fu, Y., Colonna, V., Mu, X.J., Kang, H.M., Lappalainen, T., Sboner, A., Lochovsky, L., Chen, J., Harmanci, A., et al.; 1000 Genomes Project Consortium (2013). Integrative annotation of variants from 1092 humans: application to cancer genomics. Science *342*, 1235587.

Kim, J.H., Dhanasekaran, S.M., Prensner, J.R., Cao, X., Robinson, D., Kalyana-Sundaram, S., Huang, C., Shankar, S., Jing, X., Iyer, M., et al. (2011). Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer. Genome Res. *21*, 1028–1041.

Kluth, M., Hesse, J., Heinl, A., Krohn, A., Steurer, S., Sirma, H., Simon, R., Mayer, P.S., Schumacher, U., Grupp, K., et al. (2013). Genomic deletion of MAP3K7 at 6q12-22 is associated with early PSA recurrence in prostate cancer and absence of TMPRSS2:ERG fusions. Mod. Pathol. *26*, 975–983.

Kobayashi, Y., Absher, D.M., Gulzar, Z.G., Young, S.R., McKenney, J.K., Peehl, D.M., Brooks, J.D., Myers, R.M., and Sherlock, G. (2011). DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. Genome Res. *21*, 1017–1027.

Kunju, L.P., Carskadon, S., Siddiqui, J., Tomlins, S.A., Chinnaiyan, A.M., and Palanisamy, N. (2014). Novel RNA hybridization method for the in situ detection of ETV1, ETV4, and ETV5 gene fusions in prostate cancer. Appl. Immunohistochem. Mol. Morphol. *22*, e32–e40.

Lalonde, E., Ishkanian, A.S., Sykes, J., Fraser, M., Ross-Adams, H., Erho, N., Dunning, M.J., Halim, S., Lamb, A.D., Moon, N.C., et al. (2014). Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. Lancet Oncol. *15*, 1521–1532.

Lapointe, J., Li, C., Giacomini, C.P., Salari, K., Huang, S., Wang, P., Ferrari, M., Hernandez-Boussard, T., Brooks, J.D., and Pollack, J.R. (2007). Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis. Cancer Res. *67*, 8504–8510.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K., and Rosenfeld, M.G. (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell *139*, 1069–1083.

Lindberg, J., Kristiansen, A., Wiklund, P., Grönberg, H., and Egevad, L. (2015). Tracking the origin of metastatic prostate cancer. Eur. Urol. *67*, 819–822.

Mahapatra, S., Klee, E.W., Young, C.Y., Sun, Z., Jimenez, R.E., Klee, G.G., Tindall, D.J., and Donkena, K.V. (2012). Global methylation profiling for risk prediction of prostate cancer. Clin. Cancer Res. *18*, 2882–2895.

Mani, R.S., Tomlins, S.A., Callahan, K., Ghosh, A., Nyati, M.K., Varambally, S., Palanisamy, N., and Chinnaiyan, A.M. (2009). Induced chromosomal proximity and gene fusions in prostate cancer. Science *326*, 1230.

Mardis, E.R., Ding, L., Dooling, D.J., Larson, D.E., McLellan, M.D., Chen, K., Koboldt, D.C., Fulton, R.S., Delehaunty, K.D., McGrath, S.D., et al. (2009). Recurring mutations found by sequencing an acute myeloid leukemia genome. N. Engl. J. Med. *361*, 1058–1066.

Martin, S.T., Matsubayashi, H., Rogers, C.D., Philips, J., Couch, F.J., Brune, K., Yeo, C.J., Kern, S.E., Hruban, R.H., and Goggins, M. (2005). Increased prevalence of the BRCA2 polymorphic stop codon K3326X among individuals with familial pancreatic cancer. Oncogene *24*, 3652–3656.

Mateo, J., Hall, E., Sandhu, S., Omlin, A.G., Miranda, S., Carreira, S., Goodall, J., Gillman, A., Mossop, H., Ralph, C., et al. (2014). LBA20 - Antitumour activity of the PARP inhibitor olaparib in unselected sporadic castration-resistant prostate cancer (CRPC) in the TOPARP trial. Ann. Oncol. *25*, 1–41.

Mostaghel, E.A., Geng, L., Holcomb, I., Coleman, I.M., Lucas, J., True, L.D., and Nelson, P.S. (2010). Variability in the androgen response of prostate epithelium to 5alpha-reductase inhibition: implications for prostate cancer chemoprevention. Cancer Res. *70*, 1286–1295.

Noushmehr, H., Weisenberger, D.J., Diefes, K., Phillips, H.S., Pujara, K., Berman, B.P., Pan, F., Pelloski, C.E., Sulman, E.P., Bhat, K.P., et al.; Cancer Genome Atlas Research Network (2010). Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell *17*, 510–522.

Palanisamy, N., Ateeq, B., Kalyana-Sundaram, S., Pflueger, D., Ramnarayanan, K., Shankar, S., Han, B., Cao, Q., Cao, X., Suleman, K., et al. (2010). Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. Nat. Med. *16*, 793–798.

Parikh, C., Janakiraman, V., Wu, W.I., Foo, C.K., Kljavin, N.M., Chaudhuri, S., Stawiski, E., Lee, B., Lin, J., Li, H., et al. (2012). Disruption of PH-kinase domain interactions leads to oncogenic activation of AKT in human cancers. Proc. Natl. Acad. Sci. USA *109*, 19368–19373.

Paris, P.L., Andaya, A., Fridlyand, J., Jain, A.N., Weinberg, V., Kowbel, D., Brebner, J.H., Simko, J., Watson, J.E., Volik, S., et al. (2004). Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. Hum. Mol. Genet. *13*, 1303–1313.

Penney, K.L., Stampfer, M.J., Jahn, J.L., Sinnott, J.A., Flavin, R., Rider, J.R., Finn, S., Giovannucci, E., Sesso, H.D., Loda, M., et al. (2013). Gleason grade progression is uncommon. Cancer Res. *73*, 5163–5168.

Pettersson, A., Graff, R.E., Bauer, S.R., Pitt, M.J., Lis, R.T., Stack, E.C., Martin, N.E., Kunz, L., Penney, K.L., Ligon, A.H., et al. (2012). The TMPRSS2:ERG rearrangement, ERG expression, and prostate cancer outcomes: a cohort study and meta-analysis. Cancer Epidemiol. Biomarkers Prev. *21*, 1497–1509.

Pflueger, D., Terry, S., Sboner, A., Habegger, L., Esgueva, R., Lin, P.C., Svensson, M.A., Kitabayashi, N., Moss, B.J., MacDonald, T.Y., et al. (2011). Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing. Genome Res. *21*, 56–67.

Prandi, D., Baca, S.C., Romanel, A., Barbieri, C.E., Mosquera, J.M., Fontugne, J., Beltran, H., Sboner, A., Garraway, L.A., Rubin, M.A., and Demichelis, F. (2014). Unraveling the clonal hierarchy of somatic genomic aberrations. Genome Biol. *15*, 439.

Pritchard, C.C., Morrissey, C., Kumar, A., Zhang, X., Smith, C., Coleman, I., Salipante, S.J., Milbank, J., Yu, M., Grady, W.M., et al. (2014). Complex MSH2 and MSH6 mutations in hypermutated microsatellite unstable advanced prostate cancer. Nat. Commun. *5*, 4988.

Quon, G., Haider, S., Deshwar, A.G., Cui, A., Boutros, P.C., and Morris, Q. (2013). Computational purification of individual tumor gene expression profiles leads to significant improvements in prognostic prediction. Genome Med. *5*, 29.

Robinson, D., Van Allen, E.M., Wu, Y.M., Schultz, N., Lonigro, R.J., Mosquera, J.M., Montgomery, B., Taplin, M.E., Pritchard, C.C., Attard, G., et al. (2015). Integrative clinical genomics of advanced prostate cancer. Cell *161*, 1215–1228.

Rodrigues, L.U., Rider, L., Nieto, C., Romero, L., Karimpour-Fard, A., Loda, M., Lucia, M.S., Wu, M., Shi, L., Cimic, A., et al. (2015). Coordinate loss of MAP3K7 and CHD1 promotes aggressive prostate cancer. Cancer Res. *75*, 1021–1034.

Rohle, D., Popovici-Muller, J., Palaskas, N., Turcan, S., Grommes, C., Campos, C., Tsoi, J., Clark, O., Oldrini, B., Komisopoulou, E., et al. (2013). An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. Science *340*, 626–630.

Sboner, A., Habegger, L., Pflueger, D., Terry, S., Chen, D.Z., Rozowsky, J.S., Tewari, A.K., Kitabayashi, N., Moss, B.J., Chee, M.S., et al. (2010). FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. Genome Biol. *11*, R104.

Schaeffer, E.M., Marchionni, L., Huang, Z., Simons, B., Blackman, A., Yu, W., Parmigiani, G., and Berman, D.M. (2008). Androgen-induced programs for prostate epithelial growth and invasion arise in embryogenesis and are reactivated in cancer. Oncogene *27*, 7180–7191.

Schwartz, S., Wongvipat, J., Trigwell, C.B., Hancox, U., Carver, B.S., Rodrik-Outmezguine, V., Will, M., Yellen, P., de Stanchina, E., Baselga, J., et al. (2015). Feedback suppression of PI3Kα signaling in PTEN-mutated tumors is relieved by selective inhibition of PI3Kβ. Cancer Cell *27*, 109–122.

Shen, R., Olshen, A.B., and Ladanyi, M. (2009). Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. Bioinformatics *25*, 2906–2912.

Siegel, R.L., Miller, K.D., and Jemal, A. (2015). Cancer statistics, 2015. CA Cancer J. Clin. *65*, 5–29.

Taylor, B.S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B.S., Arora, V.K., Kaushik, P., Cerami, E., Reva, B., et al. (2010). Integrative genomic profiling of human prostate cancer. Cancer Cell *18*, 11–22.

Tirode, F., Surdez, D., Ma, X., Parker, M., Le Deley, M.C., Bahrami, A., Zhang, Z., Lapouble, E., Grossetête-Lalami, S., Rusch, M., et al.; St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project and the International Cancer Genome Consortium (2014). Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. Cancer Discov. *4*, 1342–1353.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science *310*, 644–648.

Tomlins, S.A., Laxman, B., Dhanasekaran, S.M., Helgeson, B.E., Cao, X., Morris, D.S., Menon, A., Jing, X., Cao, Q., Han, B., et al. (2007). Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature *448*, 595–599.

Tomlins, S.A., Bjartell, A., Chinnaiyan, A.M., Jenster, G., Nam, R.K., Rubin, M.A., and Schalken, J.A. (2009). ETS gene fusions in prostate cancer: from discovery to daily clinical practice. Eur. Urol. *56*, 275–286.

Tomlins, S.A., Alshalalfa, M., Davicioni, E., Erho, N., Yousefi, K., Zhao, S., Haddad, Z., Den, R.B., Dicker, A.P., Trock, B.J., et al. (2015). Characterization of 1577 primary prostate cancers reveals novel biological and clinicopathologic insights into molecular subtypes. Eur. Urol. *68*, 555–567.

van Dekken, H., Paris, P.L., Albertson, D.G., Alers, J.C., Andaya, A., Kowbel, D., van der Kwast, T.H., Pinkel, D., Schröder, F.H., Vissers, K.J., et al. (2004). Evaluation of genetic patterns in different tumor areas of intermediate-grade prostatic adenocarcinomas by high-resolution genomic array analysis. Genes Chromosomes Cancer *39*, 249–256.

Wang, K., Singh, D., Zeng, Z., Coleman, S.J., Huang, Y., Savich, G.L., He, X., Mieczkowski, P., Grimm, S.A., Perou, C.M., et al. (2010). MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. *38*, e178.

Wang, X.S., Shankar, S., Dhanasekaran, S.M., Ateeq, B., Sasaki, A.T., Jing, X., Robinson, D., Cao, Q., Prensner, J.R., Yocum, A.K., et al. (2011). Characterization of KRAS rearrangements in metastatic prostate cancer. Cancer Discov. *1*, 35–43.

Watson, P.A., Chen, Y.F., Balbas, M.D., Wongvipat, J., Socci, N.D., Viale, A., Kim, K., and Sawyers, C.L. (2010). Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor. Proc. Natl. Acad. Sci. USA *107*, 16759–16765.

Cell

No Evidence for Cardiomyocyte Number Expansion in Preadolescent Mice

Graphical Abstract

Highlights

- Cardiomyocyte number expansion is limited to the neonatal period

- Multinucleation and ploidisation follow neonatal cardiomyocyte duplication

- Cardiomyocyte hypertrophy is sufficient to explain preadolescent heart growth

Authors

Kanar Alkass, Joni Panula, Mattias Westman, Ting-Di Wu, Jean-Luc Guerquin-Kern, Olaf Bergmann

Correspondence

olaf.bergmann@ki.se

In Brief

Contrary to a recent report suggesting that a preadolescent burst of cardiomyocyte proliferation promotes heart growth, cardiomyocyte number expansion appears limited to the neonatal period, with cardiomyocyte hypertrophy likely accounting for the increase in the heart size.

CellPress

No Evidence for Cardiomyocyte Number Expansion in Preadolescent Mice

Kanar Alkass,[1,2] Joni Panula,[1] Mattias Westman,[3] Ting-Di Wu,[4,5] Jean-Luc Guerquin-Kern,[4,5] and Olaf Bergmann[1,*]

[1]Cell and Molecular Biology, Karolinska Institutet, SE-17177 Stockholm, Sweden
[2]Department of Forensic Medicine, The National Board of Forensic Medicine, SE-17177 Stockholm, Sweden
[3]Department of Medicine, Karolinska Institutet, SE-14186 Huddinge, Sweden
[4]Institut Curie, Centre de Recherche, F-91405 Orsay, France
[5]INSERM, U1196; CNRS, UMR9187, F-91405 Orsay, France
*Correspondence: olaf.bergmann@ki.se
http://dx.doi.org/10.1016/j.cell.2015.10.035

SUMMARY

The magnitude of cardiomyocyte generation in the adult heart has been heavily debated. A recent report suggests that during mouse preadolescence, cardiomyocyte proliferation leads to a 40% increase in the number of cardiomyocytes. Such an expansion would change our understanding of heart growth and have far-reaching implications for cardiac regeneration. Here, using design-based stereology, we found that cardiomyocyte proliferation accounted for 30% of postnatal DNA synthesis; however, we were unable to detect any changes in cardiomyocyte number after postnatal day 11. ^{15}N-thymidine and BrdU analyses provided no evidence for a proliferative peak in preadolescent mice. By contrast, cardiomyocyte multinucleation comprises 57% of postnatal DNA synthesis, followed by cardiomyocyte nuclear polyploidisation, contributing with 13% to DNA synthesis within the second and third postnatal weeks. We conclude that the majority of cardiomyocytes is set within the first postnatal week and that this event is followed by two waves of non-replicative DNA synthesis. This Matters Arising paper is in response to Naqvi et al. (2014), published in *Cell*. See also the associated Correspondence by Soonpaa et al. (2015), and the response by Naqvi et al. (2015), published in this issue.

INTRODUCTION

The replacement of cardiomyocytes has been a major challenge in regenerative medicine. The neonatal mouse heart exhibits robust myogenesis after apical resection and ischemic lesions; this myogenesis is mainly mediated by duplication of preexisting cardiomyocytes (Ali et al., 2014; Porrello et al., 2011; Puente et al., 2014), although there is also evidence for a contribution of precursor cells (c-kit-positive) to the regenerating myocardium (Jesty et al., 2012). This regenerative process seems to be limited to the first postnatal week (Puente et al., 2014), which

coincides with an increase in binucleation in cardiomyocytes (Soonpaa et al., 1996; Walsh et al., 2010). Senyo et al. evaluated the degree of cardiomyocyte proliferation via the detection of the non-radioactive isotope ^{15}N-thymidine in dividing cardiomyocytes. They observed limited myogenesis in both young and old mice, with annual proliferation rates of less than 1% (Senyo et al., 2013), in agreement with the human cardiomyocyte renewal rates established by ^{14}C birth dating (Bergmann et al., 2009; 2015)

In contrast to these findings, Naqvi et al. (2014) reported a second wave of cardiomyocyte proliferation during preadolescence that occurs in a highly synchronized fashion. Most cardiomyocytes re-entered the cell cycle starting on the evening of postnatal day 14, followed by mitosis and cytokinesis on postnatal day 15. This proliferative event remarkably increased the cardiomyocyte count by 40%. As most cardiomyocytes are binucleated at this stage, the authors suggested a model in which binucleated cardiomyocytes undergo karyokinesis, resulting in tetranucleated cardiomyocytes, followed by cytokinesis to generate two mononucleated cardiomyocytes and one binucleated cardiomyocyte. These replicative events were suggested to be mediated by thyroid hormone (T3) through the IGF-1/Akt pathway. These observations were provocative because another recent study demonstrated that the increase in oxygen at birth induces DNA damage, leading to cell-cycle arrest in cardiomyocytes shortly after birth (Puente et al., 2014). The findings reported by Naqvi et al. suggest a more complex regulation of cardiomyocyte proliferation and consequently have major implications for the understanding of cardiomyocyte proliferation (Palpant and Murry, 2014; Zhang and Kühn, 2014). We have re-examined their key observations, using both similar and alternative approaches. We observed that cardiomyocyte expansion is restricted mainly to the first postnatal week, seriously challenging the reported contribution of cardiomyocyte proliferation to the growing preadolescent mouse heart. Instead, we observed that, in addition to multinucleation, murine cardiomyocytes undergo polyploidization in the second and third postnatal weeks. This time period corresponds well to the increase in ploidy in human preadolescent hearts (Bergmann et al., 2009; Mollova et al., 2013). This study also supports recent reports of substantial early postnatal cardiomyogenesis without finding any evidence for a second wave of cardiomyocyte proliferation contributing to heart growth.
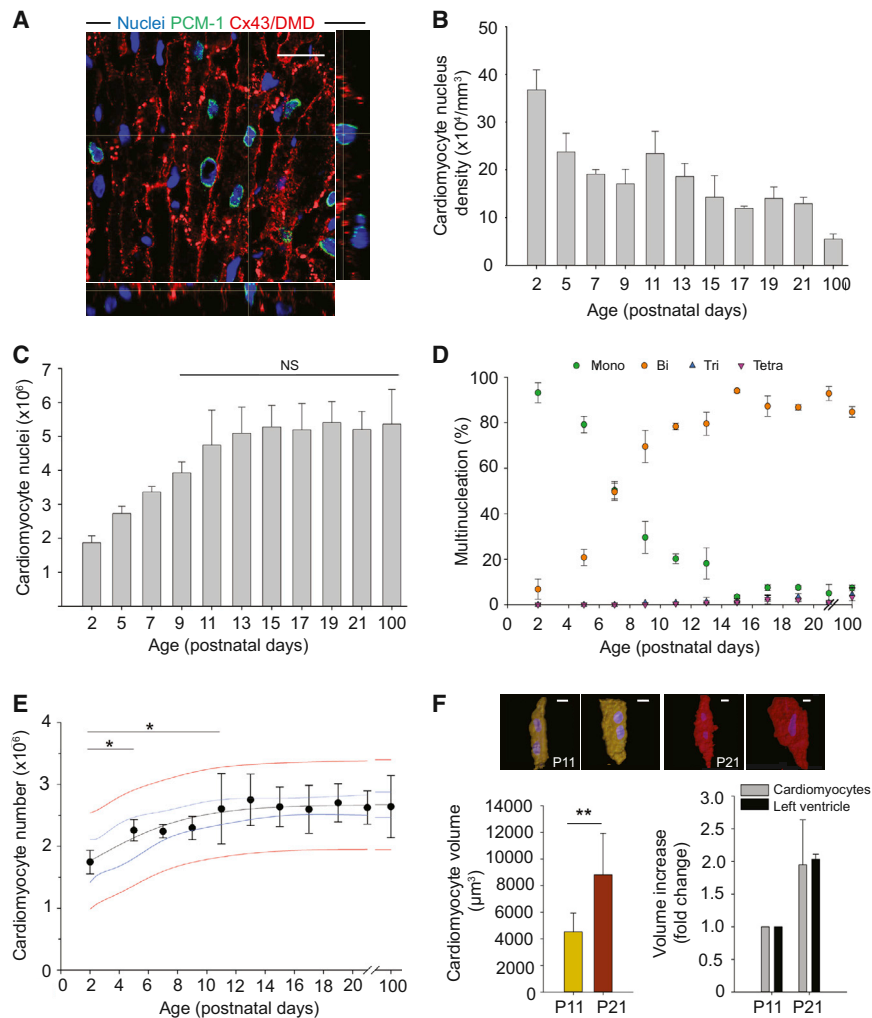
Figure 1. Expansion of the Pool of Cardiomyocytes in Neonatal Mouse Hearts

(A) The labeling strategy used to unequivocally identify cardiomyocyte nuclei and the number of nuclei per cardiomyocyte. Cardiomyocyte nuclei were labeled with antibodies against PCM-1 (green), and the cardiomyocyte cell borders were labeled with antibodies against connexin43 (Cx43) and dystrophin (DMD) (red). The scale bar indicates 20 μm.

(B) The cardiomyocyte density decreased during heart growth from P2 to P100.

(C) The number of cardiomyocyte nuclei plateaued around P11 and remained constant thereafter (n = 3 to 6 for all analyzed time points).

(D) The process of multinucleation began early in the neonatal period. At P9, most cardiomyocytes already contained two or more nuclei.

(E) Stereological analysis revealed that the number of cardiomyocytes increased in the early neonatal period from 1.7 × 10⁶ on P2 to 2.26 × 10⁶ on P5 and reach a plateau on P11 (95% confidence interval (blue), 95% prediction interval (red)).

(F) The results of the volume analysis of isolated cardiomyocytes on P11 and P21. The upper panel shows cardiomyocytes isolated on P11 (yellow) and P21 (red) rendered by the Imaris software to obtain their individual volumes (n = 3 for both groups, total of 274 cardiomyocytes analyzed). Scale bars,10 μm. Note the different lengths of the scale bars in the images of cardiomyocytes from P11 and P21. The left ventricle (black bars) increased 2.0-fold between P11 and P21, comparable to the increase in the average cardiomyocyte volume (gray bars) (right). * indicates p < 0.05; ** indicates p < 0.001; NS: not significant. All error bars indicate SD.

RESULTS

The Final Number of Cardiomyocytes Is Mainly Established within the First Postnatal Week

We determined the number of cardiomyocytes by stereology in mouse hearts from postnatal day 2 (P2) to postnatal day 100 (P100) (Figure 1). The identification of cardiomyocyte nuclei in tissue sections is challenging, particularly in the perinatal period when the cell density is very high (Ang et al., 2010; Soonpaa and Field, 1998). To circumvent this problem, we used antibodies against the cardiomyocyte nuclear marker pericentriolar material 1 (PCM-1) (Figure 1A) (Bergmann and Jovinge, 2012; Bergmann et al., 2011; Gilsbach et al., 2014). We obtained the mass of the left ventricle by weighting the left ventricle including the septum (Figures S1A and S1B). The reference volume was calculated using the tissue density of the myocardium (1.06 g/cm³) (Brüel and Nyengaard, 2005). The density of myocyte nuclei gradually decreased from 367,504 ± 42,055/mm³ (mean ± SD) on P2 to 128,983 ± 13,555/mm³ on P20 and 55,085 ± 10,574/mm³ on P100 (Figure 1B).

Next, we established the number of postnatal cardiomyocyte nuclei (Experimental Procedures). The number of cardiomyocyte nuclei expanded continuously between P2 (1.87 × 10⁶ ± 0.20 × 10⁶) and P11 and plateaued thereafter (4.75 × 10⁶ ± 1.04) (ANOVA, post hoc Holm-Sidak, p > 0.05) (Figure 1C). The ratio of mono- to multinucleated cardiomyocytes changed substantially during the first 10 postnatal days. On P2, the majority (93.2% ± 4.4%) of cardiomyocytes remained mononucleated, in agreement with previous studies, whereas on P11, only 20.2% ± 2.2% were mononucleated, and 78.4% ± 1.5% were binucleated. We did not observe any further changes in this ratio between P11 and P100 (Figure 1D). Taking multinucleation into account, we established that cardiomyocytes expanded by approximately 40% between P2 (1.7 × 10⁶ ± 0.2 × 10⁶) and P5 (2.3 × 10⁶ ± 0.2 × 10⁶) (ANOVA, post hoc t test with Holm-Bonferroni correction, p < 0.05) and then plateaued on P11 (2.6 × 10⁶ ± 0.6 × 10⁶) and remained constant at least until P100 (linear regression, R = 0.016, p = 0.935, Figures 1E and S1C).

To investigate the contribution of the increase in cardiomyocyte volume to the preadolescent growth of the left ventricle (Leu et al., 2001), we determined the average volume of
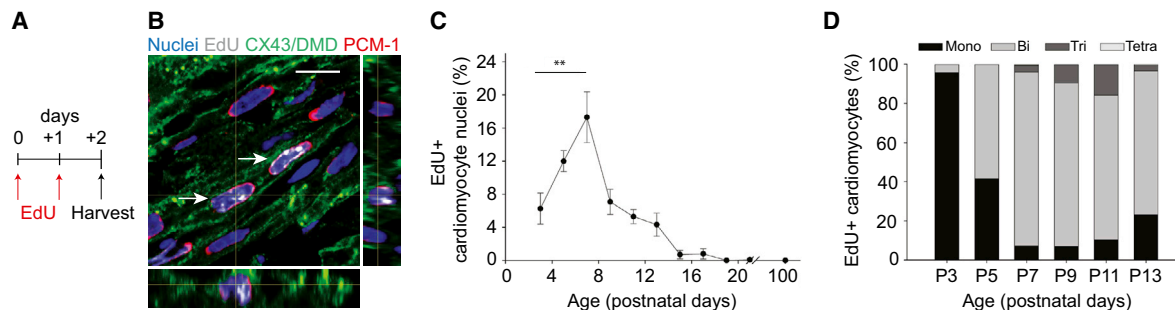
Figure 2. Cardiomyocyte DNA Synthesis in the Growing Mouse Heart
(A) Two EdU pulses were given on 2 consecutive days. Mice were sacrificed after a 24-hr chase period.
(B) EdU incorporation was detected in cardiomyocyte nuclei by co-labeling with PCM-1, and connexin43 (Cx43) and dystrophin (DMD) staining were used to delineate cell borders. Arrows indicate EdU-positive cardiomyocyte nuclei. The scale bars, 10 μm.
(C) The postnatal DNA synthesis in cardiomyocytes was highest in mice sacrificed on P7 (∼17%), and rapidly declined thereafter to values below 1% on P15 (n = 3–4, for all analyzed time points).
(D) Mononucleated cardiomyocytes incorporated most of the EdU on P3, but by P7, most cardiomyocytes were bi- or multinucleated. The relative increase in the number of mononucleated EdU-positive cardiomyocytes on P13 may be related to the increase in nuclear ploidy (Figure 3). ** indicates p < 0.001.
All error bars indicate SD.

cardiomyocytes on day P11 (4,530 μm^3 ± 1,410 μm^3) and P21 (8,820 μm^3 ± 3,120 μm^3) (t test, p < 0.01) (Figure 1F and Experimental Procedures). We determined that the 2.0-fold increase in the volume of the left ventricle (P11: 20.1 ± 1.3 mm^3 and P21: 40.8 ± 1.6 mm^3) could be fully explained by the volume increase (2.0-fold) of cardiomyocytes, indicating mainly hypertrophic growth of the left ventricle and establishment of the full complement of cardiomyocytes by P11.

Postnatal DNA Synthesis Changes Mononucleated Cardiomyocytes into Multinucleated Cardiomyocytes
The mice were treated with the thymidine analog EdU on two consecutive days and sacrificed one day after the second treatment (Figure 2A). The frequency of EdU incorporation was determined by co-labeling with antibodies against cardiomyocyte nuclei (PCM-1) (Figures 2B and C). The EdU labeling frequency peaked on P7 (17.3% ± 3.1%) (ANOVA, p < 0.001, post hoc, Holm-Sidak, p < 0.001), followed by a continuous decline to ratios of less than 1% on P15. On P100, we identified only one EdU-labeled cardiomyocyte nucleus among 980 analyzed nuclei (Figure 2C). Next, we investigated the distribution of EdU incorporation into mono- and multinucleated cardiomyocytes (Figures 2B and 2D). In agreement with cardiomyocyte proliferation, EdU incorporation was predominantly observed in mononucleated cardiomyocytes on P3 (95.8% ± 1.9%) and rapidly changed during the first postnatal week (P7), when most EdU-labeled cardiomyocytes became binucleated (mononucleated: 7.3% ± 3.5%, binucleated: 88.7% ± 5.2%, and multinucleated: 4.0% ± 2.2%) (Figure 2D). Our data support a previous report suggesting an early postnatal switch from cardiomyocyte proliferation to multinucleation (Soonpaa et al., 1996; Walsh et al., 2010).

Cardiomyocyte DNA Content Increases during the Second and Third Postnatal Weeks
Many cardiomyocyte nuclei undergo DNA synthesis to become polyploid in humans (Adler, 1991; Bergmann et al., 2009; Herget et al., 1997; Mollova et al., 2013). However, ploidy changes in juvenile mouse hearts have not been systematically investigated. We therefore determined the DNA content of murine cardiomyocyte nuclei on different postnatal days by flow cytometry (Figures 3A to 3D, and S2). Most cardiomyocyte nuclei were diploid around birth until P9 (Figures 3C, left, and S2), corresponding to 104.7% ± 0.4% in Figure 3D. Between postnatal weeks 2 and 3, the average DNA content per nucleus increased by approximately 10% in the left ventricle (Figures 3C, right, and S2) to 115.5% ± 2.3% (Figure 3D, ANOVA p < 0.001) and remained constant thereafter until P100 (Figure 3D). This finding is also reflected by an increase in the ploidy of the EdU-labeled cardiomyocyte nuclei between P3 and P13 (Figures 3E and 3F).

Cardiomyocytes Exhibit No Enhanced Cell-Cycle Activity in Preadolescent Mice
Ki-67 is expressed in all phases of the cell cycle including mitotic events (Scholzen and Gerdes, 2000). Thus, we used Ki-67 to exclude substantial cell-cycle activity on P14 and P15 as described by Naqvi et al. (2014). We sacrificed postnatal mice on P14 at 9:00 p.m., on P15 at noon, and on P16 at 9:00 a.m. The shorter length of our analysis intervals (14 hr and 21 hr) compared to the average mammalian cell-cycle length of approximately 24 hr (Hahn et al., 2009; Ponti et al., 2013) would allow us to detect any cell-cycle activity on P15. However, we could not detect any significant difference between P14, P15, and P16 (Holm-Sidak method, p > 0.05). By contrast, we observed that these three time points all exhibited Ki-67 frequencies of less than 2% compared to that on P7 (ANOVA, p < 0.001) (Figures 4A and 4B).

PCM-1 disassembles in pro-metaphase and metaphase of mitosis (Srsen et al., 2009). Although PCM-1 can be detected in all other phases of the cell cycle, we excluded the possibility that we underestimated the number of cycling cardiomyocytes by also using antibodies against cardiac troponin I to identify cardiomyocyte nuclei by their location in the cardiomyocyte cytoplasm (Figure 4B).
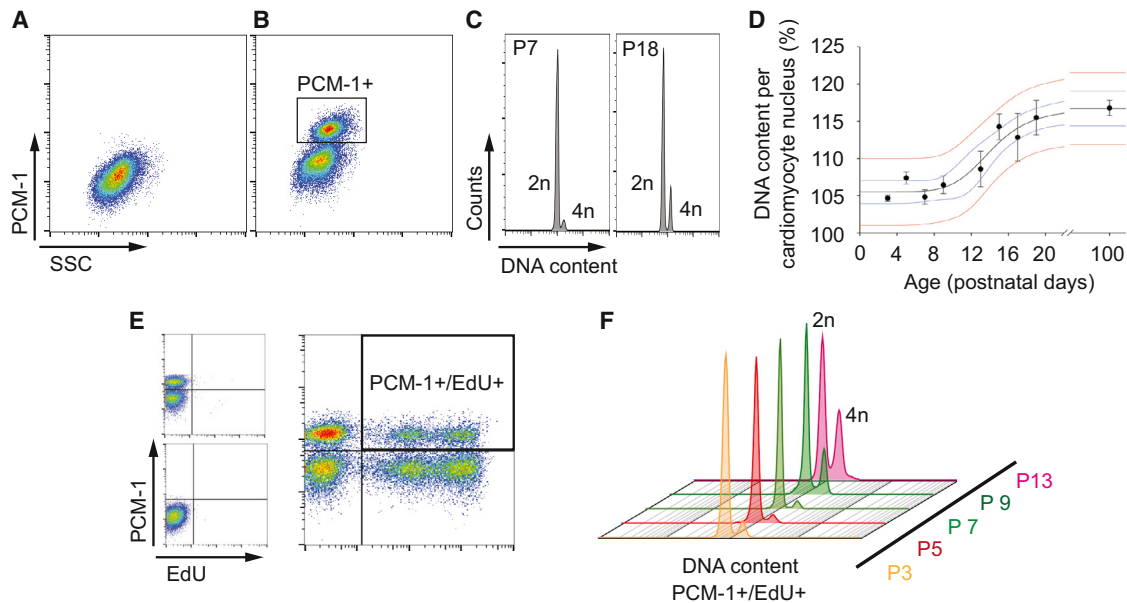
Figure 3. Cardiomyocyte Nuclear Ploidy in the Growing Mouse Heart

(A and B) Cardiomyocyte nuclei were labeled with isotype controls (A) and antibodies against PCM-1 (B).

(C) Two representative flow cytometry histograms depict the DNA content of cardiomyocyte nuclei (PCM-1-positive) from P7 and P18 mice.

(D) Time course of the cardiomyocyte DNA content per nucleus. Cardiomyocyte nuclei undergo polyploidization during the second and third postnatal weeks (95% confidence interval [blue], 95% prediction interval [red]; 100% corresponds to a pure diploid population [2n], and 200% to a pure tetraploid population [4n]).

(E and F) Flow cytometric analysis of the EdU-positive cardiomyocyte nuclear ploidy (see Figure 2A for the time course of EdU injection). (E) Representative flow cytometry plots for immunolabeling of P7 animals with antibodies against PCM-1 and EdU detection (PCM-1+/EdU+). Note that the different fluorescent intensities within the EdU-positive populations indicate different levels of EdU incorporation. Cells may have gone through the S-phase twice after BrdU was delivered. Lower left: flow cytometry plots with isotype control antibodies and without EdU detection; upper left: flow cytometry plots with PCM-1 antibodies and without EdU detection.

(F) The nuclear ploidy of EdU-positive cardiomyocyte nuclei measured at different time points (P3 to P13). The nuclear ploidy began to increase on P9. 2n: diploid; 4n: tetraploid; SSC: side scatter.

All error bars indicate SD.

BrdU Incorporation Reveals Only Minimal DNA Synthesis in Preadolescent Cardiomyocytes

Naqvi et al. proposed a highly synchronized last round of cardiomyocyte proliferation beginning on the evening of P14. Thymidine analogs such as BrdU have a short biological half-life. Therefore, it might be difficult to detect the suggested cardiomyocyte DNA synthesis on P15 with a single BrdU injection. Consequently, we subcutaneously implanted a pellet on P13 that continuously released BrdU (Experimental Procedures and Figures 4C to 4I). BrdU incorporation into cardiomyocyte nuclei (PCM-1-positive) was detected by immunohistochemistry (IHC) (Figures 4D and 4E) and flow cytometry on P18 (Figures 4F to 4H). Both experimental strategies revealed a BrdU incorporation rate of less than 3% in cardiomyocyte nuclei (IHC: 1.8% ± 0.9%; flow cytometry: 2.9% ± 1.4%) (Figure 4I). An analysis of the DNA content in BrdU-positive nuclei revealed an increase in nuclear ploidy in cardiomyocytes compared to non-cardiomyocytes (Figure 4H). These findings exclude the possibility of a major proliferative cardiomyocyte burst between P13 and P18. By contrast, the majority of BrdU-positive cardiomyocyte nuclei were tetraploid, indicating that DNA synthesis is linked to polyploidization rather than proliferation after P13 (also see Figure 3). Furthermore, we did not observe any differences in the BrdU labeling frequency of endocardial (1.4% ± 0.4%) or epicardial car-

diomyocyte (1.3% ± 0.6%) nuclei (paired t test, p = 0.93) (Figures 4D and S3).

Multi-Isotope Mass Spectrometry Analyses of 15N Thymidine Incorporation Provide No Evidence of a Second Wave of Cardiomyocyte Proliferation in Preadolescent Mice

We continuously administered 15N-thymidine (Experimental Procedures) in preadolescent mouse hearts (P13–P23) (Figure 5A) to exclude the possibility that continuous delivery of the halogenated thymidine analogs (BrdU or EdU) might be associated with toxic effects that alter cell turnover (Andersen et al., 2013; Wilson et al., 2008). Nonradioactive stable isotope tracers such as 15N-thymidine do not alter biochemical reactions and are not harmful to the animal (Steinhauser et al., 2012). Multi-isotope imaging mass spectrometry (MIMS) allows the simultaneous detection of the stable isotopes of the same element (Senyo et al., 2013). Due to the low natural abundance of 15N (0.37%), the incorporation of a 15N-labeled tracer is readily detected due to an increase in the 15N:14N ratio. Because DNA contains a high amount of phosphate, we also visualized phosphorus (31P) in addition to the 15N:14N ratio to identify cardiac nuclei with high spatial resolution (Figures 5B to 5D). At a lateral resolution of less than 100 nm,

cardiomyocytes were identified based on the cell borders and their subcellular specific ultrastructure, allowing the identification of cardiomyocyte nuclei based on their location (Figures 5B to 5D). The nuclear 15N integration was evenly distributed throughout the investigated myocardium (Figure S4). In addition to analyzing 15N thymidine incorporation on sections, we isolated cardiomyocyte nuclei by flow cytometry (as shown in Figure 3A and 3B), and determined 15N thymidine incorporation subsequently (Figure 5E). In both experimental designs we observed 15N-thymidine incorporation in only a small fraction of cardiomyocyte nuclei between P13 and P23 (0.9% to 2.1%) (t test, p > 0.05) (Figures 5B to 5F), comparable to the data obtained by BrdU infusion (Figures 4C to 4I) and not compatible with the robust cardiomyocyte proliferation suggested by Naqvi et al.

An Integrated Model of DNA Synthesis in Postnatal Cardiomyocytes

Based on our data obtained by design-based stereology (Figures 1A to 1E) and by flow cytometric analysis (Figure 3D), we established a quantitative model of DNA synthesis in postnatal cardiomyocytes (Figures 6A to 6C). Cardiomyocyte proliferation is highest at birth, followed by a phase of multinucleation reaching a maximum around P7. The polyploidization of cardiomyocyte nuclei reflects the last wave of DNA synthesis, peaking around P14 (Figure 6B). Derivative graphs depicting the DNA content changes per time unit were calculated based on absolute quantifications of cardiomyocyte nucleus number (Figure 1C), cardiomyocyte ploidy (Figure 3D), and cardiomyocyte number (Figure 1E) without relying on markers of proliferation or the incorporation of thymidine analogs with unknown biological half-lives. As the number of cardiomyocytes and number of cardiomyocyte nuclei substantially changes postnataly, we used the time of birth as a reference for calculating postnatal DNA synthesis. Multinucleation accounted for 57% of the total postnatal DNA synthesis, cardiomyocyte number expansion was related to 30%, and polyploidization reflected 13% of this synthesis, mainly in the second and third postnatal weeks (areas under the curves, see Figure 6).

DISCUSSION

The postnatal heart has a robust capability to generate new myocardium after apical dissection and injury, although whether induced myocardiogenesis leads to complete myocardial regeneration remains controversial (Andersen et al., 2014; Jesty et al., 2012; Mahmoud et al., 2013; Sadek et al., 2014). Until recently, whether neonatal myocardiogenesis is a physiological phenomenon or whether injury is required to initiate substantial cardiomyocyte proliferation has been unknown. We demonstrated that, even in the uninjured neonatal heart, a substantial number of cardiomyocytes are generated within the first postnatal week, in agreement with the report by Naqvi et al. (Naqvi et al., 2014). However, by several distinct approaches, we failed to observe any substantial increase in the cardiomyocyte number or proliferation rate between P13 and P100.

We first used design-based stereology to establish the cardiomyocyte cell count in the left ventricle. Stereology is a powerful tool to obtain unbiased estimates of cell and nucleus numbers in different organ systems (Bergmann et al., 2015; Yeung et al., 2014). This technique relies on the accurate identification of cardiomyocyte nuclei; we and others have successfully utilized pericentriolar material 1 (PCM-1) as a specific marker for cardiomyocyte nuclei (Bergmann and Jovinge, 2012; Bergmann et al., 2011; Bergmann et al., 2015; Gilsbach et al., 2014; Preissl et al., 2015). In contrast to cell isolation strategies, stereological estimates are not dependent on isolation efficiency. Even with the most efficient method, retrograde perfusion using a Langendorff system, the expected yield of cardiomyocytes is never 100% and varies between ∼1.5 and ∼2.5 million in mice (older than P17) (Naqvi et al., 2014). Our cardiomyocyte number estimate is supported by previous studies that reported similar number of cardiomyocyte nuclei and cardiomyocyte nucleus density in young adult mice (Adler et al., 1996; Bersell et al., 2009). Taking our stereological estimation as a reference for the cardiomyocyte count (Figure 1E), Naqvi et al. cardiomyocyte isolation efficiency is estimated to be only 50% to 65% depending on the analyzed time points (Figures S1D and S1E). Because the amount and composition of the extracellular matrix changes dramatically in growing hearts (Anderson, 2010), it is likely that the reported cardiomyocyte increase between P15 and P17 could, at least in part, be explained by age-dependent differences in the cardiomyocyte isolation efficiencies. In this case, already changes in isolation efficiency of 15%–20% could explain the discrepancy between Naqvi et al. and our data (Figures S1D and S1E).

We performed Ki-67 staining on P14, P15, and P16 and obtained no evidence of extensive cardiomyocyte proliferation (<2% labeling frequency). By contrast, we observed that up to 13% of cardiomyocytes were in the cell cycle on P7, when most multinucleation occurs. Ki-67 expression is not restricted to mitosis but can be detected in all cell-cycle stages (G1, S, G2, and mitosis). For a cell-cycle length of approximately 24 hr (Hahn et al., 2009; Ponti et al., 2013), an extensive burst of proliferation would have been readily detected by this labeling strategy. One of the major caveats of Naqvi et al. immunohistochemistry approaches is the lack of a nuclear marker to unequivocally identify cardiomyocytes as discussed above. Without using a cardiomyocyte nuclear marker, it is difficult to distinguish cardiomyocyte from non-cardiomyocytes (Ang et al., 2010), particular in growing hearts in which the nucleus density can be more than 10-times higher than in the adult hearts (Bergmann et al., 2015). It is our impression that the provided images in Naqvi et al. do not allow an accurate discrimination between cardiomyocytes and non-cardiomyocytes being aurora B-positive (see Naqvi et al., Figures 3B, S2, and S3). Moreover, Naqvi et al. reported that approximately 30% of isolated cardiomyocytes are positive for aurora B (compared to 15% measured in tissue sections). This labeling frequency even exceeds mitotic rates in embryonic hearts by far (embryonic day 14.5: 13.9% in G2/M phase) (Walsh et al., 2010). The measured high aurora B frequency of 30% in the cardiomyocyte compartment would result in an even more dramatic increase in the number of cardiomyocytes than reported by Naqvi et al. (22% based on Langendorff preparation on P15). Given the estimated short duration of mitosis (approx. 1.8 hr in ventricular cardiomyocytes) (Mollova
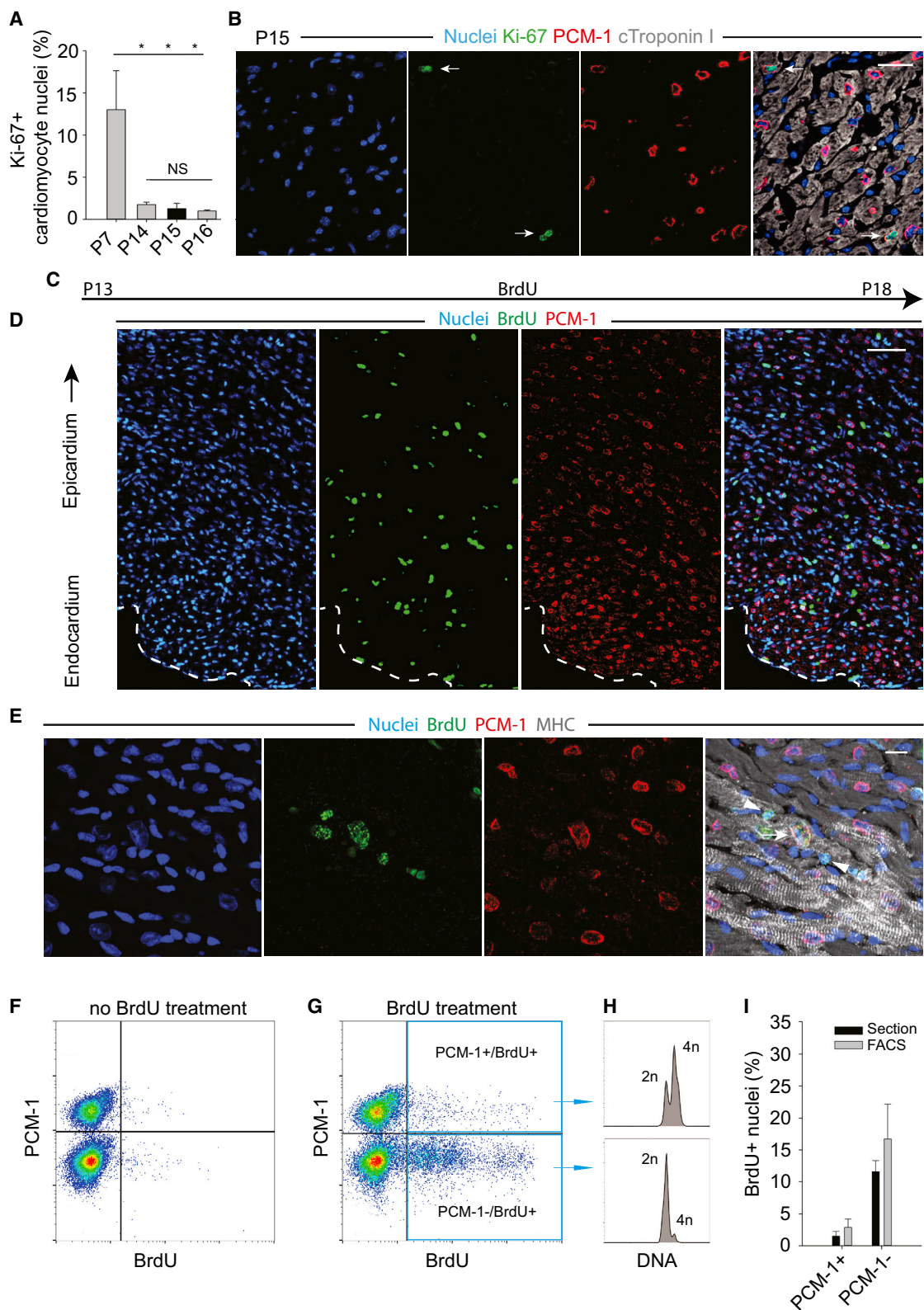
(legend on next page)

et al., 2013), 30% of mitotic cardiomyocytes in this four-hour window would translate to an expansion of the cardiomyocyte compartment by 67% (30% × 4 hr/1.8 hr). Therefore, it seems reasonable to assume that Naqvi et al. have substantially overestimated the number of aurora B-positive cardiomyocytes in their study.

One of the most prevalent techniques for detecting cell proliferation is the administration and immunohistochemical detection of thymidine analogs such as BrdU or EdU. EdU injection in the afternoon on P13 and P14 resulted in a fraction of less than 1% labeled cardiomyocyte nuclei when analyzed on P15 (Figure 2C). However, given the relatively short biological half-life of thymidine analogs, a highly synchronized proliferation of cardiomyocytes on P14 and P15 could go undetected. Therefore, we continuously administered BrdU by implanting a subcutaneous pellet on P13. In agreement with our other results, less than 3% of all cardiomyocyte nuclei had incorporated BrdU on P18 (Figure 4I), which is not compatible with the 40% increase in the cardiomyocyte cell number reported by Naqvi et al. We were also unable to confirm the regional differences in cardiomyocyte proliferation (epicardial versus endocardial regions) suggested by Naqvi et al. (Figure 4D). Furthermore, the BrdU-labeled cardiomyocyte nuclei between P13 and P18 were mainly tetraploid (Figure 4H), indicating that the majority of DNA synthesis after P13 can be attributed to polyploidization and not proliferation.

In contrast to our present study, Naqvi et al. performed a single BrdU pulse on the night of P14, which was sufficient to label a substantially larger fraction (11.3%) than detected by us of mainly diploid cardiomyocyte nuclei after a chase period of 4 days. Although, their BrdU pulse-chase experiments do not support the reported 40% increase in the cardiomyocyte cell number, their BrdU figure is still several-fold higher than in our present study. Although it is difficult to draw firm conclusions without a detailed analysis of primary data, our BrdU data obtained by both flow cytometry and by immunohistochemistry point to the possibility that Naqvi et al. cardiomyocyte nucleus isolation strategy was not rigorous enough to sufficiently remove non-cardiomyocyte nuclei from the analysis.

Because thymidine analogs (e.g., BrdU) induce the proliferation of hematopoietic stem cells (Wilson et al., 2008) and mediate toxic effects (Andersen et al., 2013), we continuously delivered biologically inert ^{15}N thymidine to detect DNA synthesis between P13 and P23 in cardiomyocytes. The labeling frequencies of cardiomyocyte nuclei were consistent with the results of our previous experiments using BrdU. Less than 3% of

all cardiomyocyte nuclei incorporated ^{15}N-thymidine during preadolescence.

In the present study, we utilized rigorous strategies including design-based stereology, cell-cycle marker analysis (Ki-67), EdU pulse chase experiments, and continuous delivery of BrdU and ^{15}N thymidine. Furthermore, we analyzed cardiomyocyte proliferation both in sections and by flow cytometry without finding any evidence for a burst of proliferation or a distinct mode of cell division between P13 and P18. Instead, we observed completion of cardiomyocyte number expansion no later than P11. We further demonstrated that young cardiomyocytes exhibit multinucleation followed by polyploidization as the cells increase in size.

Cardiomyocytes lose their ability to complete the cell cycle to increase their number. This loss leads to a progressive restriction of the cardiomyocyte cell cycle in which cytokinesis is first exchanged for multinucleation, which in turn is exchanged for polyploidization. In human hearts, the pool of cardiomyocytes and the degree of multinucleation are established soon after birth (Bergmann et al., 2015; Mayhew et al., 1997), although one report has suggested that the cardiomyocyte number increases until adulthood (Mollova et al., 2013). Polyploidization occurs in both mouse and human cardiomyocytes, to different degrees. In humans, approximately 60% of all cardiomyocyte nuclei become polyploid (Bergmann et al., 2015), whereas only 10% of all cardiomyocyte nuclei become polyploid in mice. However, polyploidization occurs after multinucleation at a similar time point in both species, mostly in preadolescence. Polyploidization occurs mainly in the second and third postnatal week. Therefore it partly overlaps with Naqvi et al. suggested increase in the number of cardiomyocytes between P15–P18. However, the limited extent of polyploidization (approximately 10%) makes it unlikely that Naqvi et al. has mistaken the moderate ploidy increase for proliferation.

Given the important role of IGF-1 signaling in cardiomyocyte growth and hypertrophy (Carrasco et al., 2014; Delaughter et al., 1999), one might speculate that the T3-induced activation of the IGF-1/IGF-1-R/Akt pathway, as suggested by Naqvi et al., mainly triggers hypertrophic cardiomyocyte growth associated with polyploidization rather than cardiomyocyte proliferation in preadolescent mice.

We used the C57bl/6N mouse strain, which is a substrain to the strain (C57bl/6J) used by Naqvi et al. (Naqvi et al., 2014). A comparison of several strains, including ICR/CD1, C3Heb/FeJ, C57bl/6BRL, and C57bl/6J showed only little variation regarding the proliferative capacity and growth in the postnatal mouse

Figure 4. No Evidence for a Peak of Cardiomyocyte Proliferation in Preadolescent Mice
(A–B) Mice were sacrificed on P7, P14 (7 pm), P15 (noon), and on P16 (9.00 am). PCM-1 (red) and cardiac troponin I (gray) labeling identified cardiomyocyte nuclei (n = 3 for all analyzed time points). Cycling cardiomyocytes were identified by Ki-67 (green) expression. Arrows indicate Ki-67-positive cardiomyocyte nuclei. The black bar depicts the Ki-67 frequency on P15. The scale bar indicates 20 μm.
(C–I) BrdU was continuously delivered via subcutaneous pellets from P13 to P18 (Experimental Procedures). BrdU incorporation into cardiomyocyte nuclei (PCM-1-positive) was detected by immunohistochemistry (D, E, and I), and by flow cytometry (F to I). (D) Tile scans of a horizontal section of the mouse heart and high-magnification images (E) revealed BrdU integration into cardiomyocyte (arrows) and non-cardiomyocyte nuclei (arrowheads). Dashed lines indicate the endocardium. Scale bars: 50 μm in (D) and 10 μm in (E). (F and G) The results of flow cytometric analyses of non-BrdU-treated mice (F) and BrdU-treated mice (G). (H) Most BrdU incorporation between P13 and P18 could be detected in tetraploid cardiomyocyte nuclei, whereas most non-cardiomyocyte nuclei were diploid. The different fluorescence intensities within the BrdU-positive populations indicate different levels of BrdU incorporation. Cells may have gone through the S-phase twice after BrdU was delivered. (I) In heart sections (n = 3) and in isolated cardiac nuclei (n = 4), less than 3% of all cardiomyocyte nuclei incorporated BrdU between P13 and P18. By contrast, ~15% of all non-cardiomyocyte nuclei incorporated BrdU. * indicates p < 0.05.
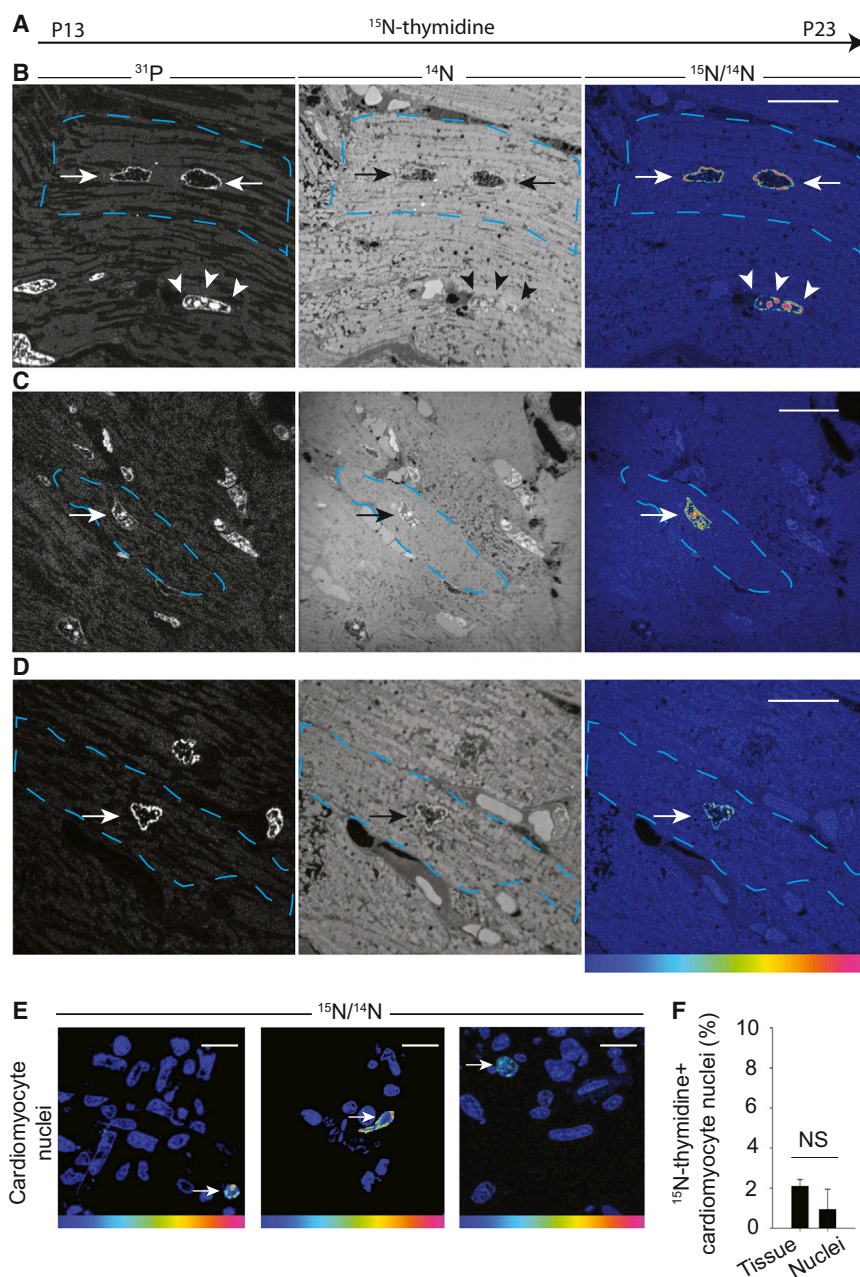All error bars indicate SD.

Figure 5. MIMS Analysis of ^{15}N Thymidine Incorporation in Cardiomyocytes of Preadolescent Mice

(A–E) ^{15}N-thymidine was continuously delivered via a subcutaneous pellet between P13 and P23 (Experimental Procedures). (B to D) Cardiac nuclei were identified by MIMS-based detection of ^{31}P, (left) and ^{14}N (middle). Images for ^{14}N were obtained on the species ^{12}C^{14}N$^-$. The ^{14}N images reveal histological and subcellular details (cell borders and organelles) and therefore permit the identification of cardiomyocyte nuclei (arrows). ^{15}N/^{14}N hue-saturation-intensity images revealed an increase in the ^{15}N/^{14}N isotope ratios above natural abundances (blue, 0% excess; red, 300% excess) and therefore indicated ^{15}N-thymidine incorporation into cardiac nuclei (arrows). (E) ^{15}N thymidine incorporation in isolated cardiomyocyte nuclei (PCM-1 positive) by flow cytometry (as shown in Figure 3A and B). Arrows indicate cardiomyocyte nuclei with ^{15}N-thymidine incorporation.

(F) Both MIMS strategies revealed that less than 2.1% of all cardiomyocyte nuclei incorporated ^{15}N-thymidine between P13 and P23 (n = 6). Images B to D are indexed in Figure S4A. Scale bars indicate 20 μm. NS: not significant.

All error bars indicate SD.

EXPERIMENTAL PROCEDURES

Animals

Whole litters of five to seven C57BL/6N mice with confirmed birth dates were used. Size variations between animals of the same age were minimal (the mean coefficient of variation was among all age groups 10%). Whenever possible, littermates were chosen. For the analysis of mice on postnatal day 100 only male animals were used. Mice received EdU (50 mg/kg) by subcutaneous or intraperitoneal daily injections. BrdU and ^{15}N-thymidine (euriso-top) were delivered continuously (0.5 mg/day corresponds to 60–70 mg/kg/day) to P13 mice by subcutaneously implanted pellets (Innovative Research of America, USA), which were in place for 5 and 10 days, respectively. All procedures were approved by the local ethics committee. Hearts were removed without perfusion, and the left ventricle including the septum was dissected. The wet weight of the heart and the ventricles was determined with fine scales (±0.1 mg).

Nuclei Isolation

Cardiac nuclei were isolated as described previously (Bergmann et al., 2015). Frozen heart tissue was dissected and homogenized in lysis buffer (0.32 M sucrose, 10 mM Tris-HCL [pH 8], 5 mM CaCl$_2$, 5 mM MgAc, 2 mM EDTA, 0.5 mM EGTA, 1 mM DTT) using a T-25 Ultra-Turrax probe homogenizer (IKA Germany) at 24,000 rpm for 10 s, followed by homogenization using a type A pestle in a 40 ml glass douncer (VWR) with eight strokes per sample. The nuclear isolates were filtered through 100-μm and 60-μm nylon mesh cell strainers (BD Bioscience) and centrifuged at 600 × g for 10 min. The pellets were dissolved in sucrose buffer (2.1 M sucrose, 10 mM Tris-HCl [pH 8], 5 mM MgAc). Finally, 10 ml of sucrose buffer was added to BSA-coated ultracentrifuge tubes and overlaid

heart (Haubner et al., 2012; Leu et al., 2001; Porrello et al., 2011; Soonpaa et al., 1996). Therefore, it is unlikely that genetic differences in the used substrains can explain the fundamental differences between our and the Naqvi et al. study.

In summary, the present study demonstrates that the mouse heart displays a robust generation of cardiomyocytes postnatally. In the uninjured neonatal mouse heart, up to 30% of all cardiomyocytes are generated even after postnatal day 2, and the full complement of cardiomyocytes (>95%) is reached after 11 days. Thus, there is a strong indication that the neonatal mouse heart harbor cues the induction of de novo myocardiogenesis to regenerate the diseased adult heart.
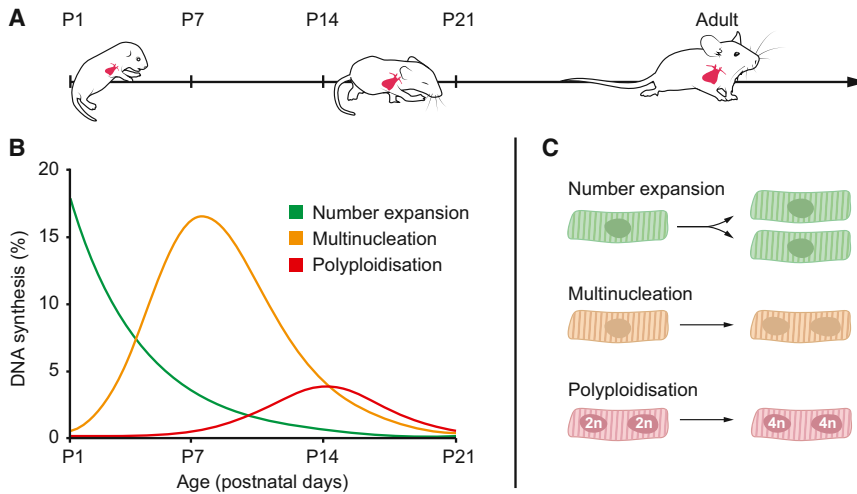
Figure 6. Cardiomyocyte DNA Synthesis in Growing Mouse Hearts

(A–C) The time course and quantification of postnatal cardiomyocyte DNA synthesis. (B and C) Cardiomyocyte proliferation, multinucleation, and polyploidization were observed in the postnatal murine heart. The magnitude and time course of the cardiomyocyte number expansion, multinucleation, and polyploidization were related. The graphs depict the DNA content changes per unit time based on stereological estimates and flow cytometric measurements of the cardiomyocyte number expansion, multinucleation, and polyploidization (Figures 1C to E and Figure 2E). The changes in DNA synthesis were related to the time around birth (P2). Diploid nuclei: 2n, tetraploid nuclei: 4n.

with the nuclear isolate. The samples were centrifuged at 13,000 × g for 60 min in a Beckman Avanti Centrifuge (Beckman Coulter), and the nuclear pellets were dissolved in NSB plus buffer (0.44 M sucrose, 10 mM Tris-HCl [pH 7.2], 70 mM KCl, 10 mM MgCl$_2$, 1.5 mM spermine). All steps were performed at 4°C.

BrdU and EdU Detection by Flow Cytometry
Nuclei were labeled with antibodies against PCM-1 (Santa Cruz, 1:200) overnight, and PCM-1 was visualized using an appropriate secondary antibody conjugated to Alexa 488 (Life Technologies). Cardiac nuclei were fixed in Fix/Perm (BD Biosciences) for 20 min on ice and sorted by flow cytometry (500,000 per tube) based on a DNA dye (Hoechst 33342). Next, the nuclei were resuspended in DNase I buffer (20 mM Tris-HCl [pH 8], 2 mM MgCl$_2$, 50 mM KCl) and incubated with 6 U of DNase I (Life Technologies) per 10^6 nuclei (37°C, 30 min). We determined that 6 U of DNase I per 10^6 nuclei is optimal for BrdU imaging of murine cardiac nuclei. The nuclei were washed once with Perm/Wash buffer (BD Biosciences) and resuspended in PBS (50 μl) for staining with BrdU antibodies conjugated to APC (BD PharMingen, 1:50, 30 min, room temperature). PCM-1 and BrdU incorporation were detected and quantified by flow cytometry. For EdU detection, we used the Click-iT EDU Alexa Fluor 647 Flow Cytometry Assay kit according to the manufacturer's instructions (Life Technologies).

MIMS Analysis
Extracted mouse tissue was minced to generate 1-mm^3 cubes and immediately fixed in 2.5% glutaraldehyde and 4% paraformaldehyde in 0.01 M PBS buffer at pH 7.2 for 2 hr at room temperature. Cardiomyocyte nuclei were isolated by flow cytometry and labeled with antibodies against PCM-1 as described in Experimental Procedures. Nuclei were transferred to 2.5% glutaraldehyde (Sigma) in 0.1 M PBS and stored in a refrigerator. Following standard electron microscopy embedding protocols, the tissue or the nuclei were postfixed in 2% OsO$_4$ in PBS for 1 hr, followed by a ddH$_2$0 wash and dehydration in an ascending alcohol series. Tissue pieces were embedded in Agar 100 resin (Agar scientific) using propylene oxide as an intermediate agent. The resin was polymerized at 60°C for 48 hr. Thin sections (150 nm) were deposited on clean silicon chips and introduced into a NanoSIMS-50 ion microprobe (CAMECA, Gennevilliers, France) operating in scanning mode (Guerquin-Kern et al., 2005). For the present study, we used a tightly focused Cs$^+$ primary ion beam, which allowed four secondary ion species (^{12}C$^-$, ^{12}C^{14}N$^-$, ^{12}C^{15}N$^-$ and ^{31}P$^-$) to be monitored in parallel from the same sputtered volume. The primary beam steps over the surface of the sample to create images of the selected ion species. After careful Cs$^+$ ion implantation to obtain steady-state ion emission, a mosaic view of the tissue over a large area was generated using a relatively high-intensity probe with a typical spot size of 200 nm (distance between 16 and 84% peak intensity from a line scan). The raster size was 80 μm

with an image definition of 256 × 256 pixels and a dwell time of 2 ms per pixel. These survey images permitted the statistical evaluation of the percentage of labeled nuclei and enabled the selection of labeled cells for further high-resolution imaging.

High-resolution images were acquired using multiframe mode. The primary beam intensity was 1 pA with a typical probe size of ~100 nm. The raster size was between 50 and 60 μm to image whole cardiomyocytes with an improved image definition of 512 × 512 pixels. With a dwell time of 2 ms per pixel, up to 20 frames were acquired, and the total analysis time was 2–3 hr. All survey images and high-resolution images were processed using ImageJ software. Before calculating local isotopic ratios, each isotopic image was properly aligned using the TOMOJ plugin (Messaoudii et al., 2007) with ^{12}C^{14}N$^-$ images used as a reference before a summed image was obtained for each ion species. A ^{15}N:^{14}N ratio map was then established from ^{12}C^{14}N and ^{12}C^{15}N images based on pixel-by-pixel division. A sample containing no labeled cells was used as a working reference to adjust the detectors prior to quantification of the ^{15}N:^{14}N ratios. The final ^{15}N:^{14}N ratio map was displayed using Hue-Saturation-Intensity (HSI) transformation. These HSI color images were generated using OpenMIMS, an ImageJ plugin developed by Claude Lechene's laboratory (Lechene et al., 2006). The hue corresponds to the ratio value, and the intensity at a given hue is an index of the statistical reliability.

Stereological Analysis
Using a design-based strategy, tissue pieces (1–2 mm diameter) from the left ventricle (including the septum) were sampled. Tissues were embedded in 8% gelatin, and isectors (spheres) with a maximum diameter of 4 mm were prepared to obtain isotropic, uniform random alignment of the samples. The isectors were embedded, and 40-μm-thick cryo sections were stained for stereological quantitation. Cardiomyocyte nuclei were stained with antibodies against PCM-1, and nuclei were stained with DRAQ5 (see Immunohistochemistry). The analysis was performed on an LSM700 confocal microscope (63× Plan-Apo oil objective) using ZEN2010b software with the NewCast Module (Visiopharm A/S, version 4.x). A minimum of 3–4 isectors were sampled, and a minimum of 200 nuclei per animal were counted (1%–2% of the area of the region of interest). A systematic random sampling scheme (meander sampling) was applied using the optical dissector with a counting frame (40 μm × 40 μm × 20 μm, and 3 μm guard zones). We defined local vertical areas where myocytes had been cut along their longitudinal axis to determine the number of nuclei per myocyte. Myocyte cell borders were labeled with connexin-43 and dystrophin. Wheat germ agglutinin (WGA) was added to facilitate the identification of the cell borders in P2 and P3 animals, and myocyte nuclei were labeled with PCM-1. To estimate the total numbers of nuclei in the heart, we utilized the two-step N$_V$ × V$_{REF}$ method (where N$_V$ is an estimate of the numerical cell density and V$_{REF}$ is the reference volume of the tissue or organ region of interest) using an optical dissector (Brüel and Nyengaard, 2005). When

necessary, the tissue shrinkage along the z axis was corrected. Tissue shrinkage along the x and y axes was not observed. The total number of cardiomyocytes was calculated based on the number of cardiomyocyte nuclei and the multinucleation level.

Volume Analysis of Isolated Cardiomyocytes

We used an adaptation of the protocol reported by Mollova et al. to assess the volume of the isolate cardiomyocytes (Mollova et al., 2013). Briefly, frozen hearts were trimmed to 1-mm^3 cubes and fixed in 4% paraformaldehyde at 4°C for 2 hr, followed by washing for 2–3 hr in HBSS buffer (Ca^{2+}, Mg^{2+}). The buffer was exchanged every hour. For cardiomyocyte isolation, collagenase B (1.8 mg/ml) and collagenase D (2.4 mg/ml) (Roche) were added to the HBSS buffer (Ca^{2+}, Mg^{2+}), and the tissue pieces were incubated for 12 to 24 hr on a slow shaker at 37°C. Isolated cardiomyocytes were concentrated by spinning at 20 × g (2 min), stained with WGA conjugated to Alexa 547 (Life Technologies, 1 mg/ml, 1:500), and mounted with ProLong Gold DAPI (Life Technologies). The entire procedure beginning with the addition of the collagenase was repeated three times. Z-stack images of cardiomyocytes were obtained using a Zeiss confocal LSM 700 microscope (63× Plan-Apo oil objective), and the individual cardiomyocyte volume was measured by the Imaris 8 (Bitplane) 3D image processing software program using the surface module.

Immunohistochemistry

Mouse hearts were incubated in sucrose solution (30% Sucrose in PBS) overnight. The tissue was embedded in Tissue Tek O.C.T. Compound (Sakura) and frozen in liquid nitrogen. After cryosectioning (−25°C), 14 μm thick sections were fixed in 2% formaldehyde for 20 min. For BrdU detection, tissue sections were incubated in 2N HCl at room temperature for 40 min prior to staining, followed by extensive washing with PBS (pH 7.4). Antibodies against BrdU (Abcam, rat 1:250), Ki67-FITC (Abcam, rabbit, SP-6, 1:100), Connexin 43 (Sigma Aldrich, rabbit, 1:10,000, and Abcam, mouse, 1:200), dystrophin (Atlas Antibodies, rabbit, 1:2,000, and Abcam, mouse, 1:200), PCM-1 (Santa Cruz, rabbit 1:200), MHC (Abcam, mouse, 1:200), and cardiac troponin I (Abcam, mouse, 1:250) were used. The antibodies were visualized using appropriate secondary antibodies conjugated to Alexa 488, 555 and/or 647 (Life Technologies, 1:500). WGA conjugated to Alexa 555 and 647 (Life Technologies, 1 mg/ml, 1:500) was used to label the cell borders in P2 and P3 animals. Nuclei were visualized with DAPI or DRAQ5 (BioStatus), and sections were mounted with ProLong medium (Life Technologies). Image analysis was performed with a Zeiss LSM 700 confocal microscope using ImageJ software.

Statistics

Significant differences between continuous variables were determined by two-tailed paired and unpaired t tests and ANOVA followed by the post hoc Holm-Sidak test or by two-tailed t tests with Holm-Bonferroni correction. The data are presented as means with SD (mean ± SD). Sigma Plot 13.0 was used for statistical analysis and for dynamic fitting of the data. Derivative graphs of the cardiomyocyte number, cardiomyocyte nuclear number and ploidy levels were generated with MathCad 15.0. DNA content changes were determined compared to the level on P2. p < 0.05 was considered significant.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.035.

AUTHOR CONTRIBUTIONS

O.B. and K.A. designed, performed, and analyzed the experiments and wrote the paper. J.P. and M.W. performed and analyzed experiments. T-D.W. and J-L.G.-K. performed the MIMS experiments. O.B. supervised the study.

ACKNOWLEDGMENTS

REFERENCES

Adler, C.P. (1991). The development and regenerative potential of cardiac muscle. In The development and regenerative potential of cardiac muscle, J.O. Oberpriller, J.C. Oberpriller, and A. Mauro, eds. (London: HAP), pp. 227–252.

Adler, C.P., Friedburg, H., Herget, G.W., Neuburger, M., and Schwalb, H. (1996). Variability of cardiomyocyte DNA content, ploidy level and nuclear number in mammalian hearts. Virchows Arch. 429, 159–164.

Ali, S.R., Hippenmeyer, S., Saadat, L.V., Luo, L., Weissman, I.L., and Ardehali, R. (2014). Existing cardiomyocytes generate cardiomyocytes at a low rate after birth in mice. Proc. Natl. Acad. Sci. USA 111, 8850–8855.

Andersen, D.C., Skovrind, I., Christensen, M.L., Jensen, C.H., and Sheikh, S.P. (2013). Stem cell survival is severely compromised by the thymidineanalog EdU (5-ethynyl-2′-deoxyuridine), an alternative to BrdU for proliferation assays and stem cell tracing. Anal. Bioanal. Chem. 405, 9585–9591.

Andersen, D.C., Ganesalingam, S., Jensen, C.H., and Sheikh, S.P. (2014). Do neonatal mouse hearts regenerate following heart apex resection? Stem Cell Reports 2, 406–413.

Anderson, P.W. (2010). Myocardium and development (Paediatric Cardiology), pp. 57–72. http://dx.doi.org/10.1016/B978-0-7020-3064-2.00004-7.

Ang, K.L., Shenje, L.T., Reuter, S., Soonpaa, M.H., Rubart, M., Field, L.J., and Galiñanes, M. (2010). Limitations of conventional approaches to identify myocyte nuclei in histologic sections of the heart. Am. J. Physiol. Cell Physiol. 298, C1603–C1609.

Bergmann, O., and Jovinge, S. (2012). Isolation of cardiomyocyte nuclei from post-mortem tissue. J. Vis. Exp. (65), 4205. http://dx.doi.org/10.3791/4205.

Bergmann, O., Bhardwaj, R.D., Bernard, S., Zdunek, S., Barnabé-Heider, F., Walsh, S., Zupicich, J., Alkass, K., Buchholz, B.A., Druid, H., et al. (2009). Evidence for cardiomyocyte renewal in humans. Science 324, 98–102.

Bergmann, O., Zdunek, S., Alkass, K., Druid, H., Bernard, S., and Frisén, J. (2011). Identification of cardiomyocyte nuclei and assessment of ploidy for the analysis of cell turnover. Exp. Cell Res. 317, 188–194.

Bergmann, O., Zdunek, S., Felker, A., Salehpour, M., Alkass, K., Bernard, S., Sjostrom, S.L., Szewczykowska, M., Jackowska, T., Dos Remedios, C., et al. (2015). Dynamics of Cell Generation and Turnover in the Human Heart. Cell 161, 1566–1575. http://dx.doi.org/10.1016/j.cell.2015.05.026.

Bersell, K., Arab, S., Haring, B., and Kühn, B. (2009). Neuregulin1/ErbB4 signaling induces cardiomyocyte proliferation and repair of heart injury. Cell 138, 257–270.

Brüel, A., and Nyengaard, J.R. (2005). Design-based stereological estimation of the total number of cardiac myocytes in histological sections. Basic Res. Cardiol. 100, 311–319.

Carrasco, L., Cea, P., Rocco, P., Peña-Oyarzún, D., Rivera-Mejias, P., Sotomayor-Flores, C., Quiroga, C., Criollo, A., Ibarra, C., Chiong, M., and Lavandero, S. (2014). Role of heterotrimeric G protein and calcium in cardiomyocyte hypertrophy induced by IGF-1. J. Cell. Biochem. 115, 712–720.

Delaughter, M.C., Taffet, G.E., Fiorotto, M.L., Entman, M.L., and Schwartz, R.J. (1999). Local insulin-like growth factor I expression induces physiologic, then pathologic, cardiac hypertrophy in transgenic mice. FASEB J. 13, 1923–1929.

Gilsbach, R., Preissl, S., Grüning, B.A., Schnick, T., Burger, L., Benes, V., Würch, A., Bönisch, U., Günther, S., Backofen, R., et al. (2014). Dynamic

DNA methylation orchestrates cardiomyocyte development, maturation and disease. Nat. Commun. 5, 5288.

Guerquin-Kern, J.L., Wu, T.D., Quintana, C., and Croisy, A. (2005). Progress in analytical imaging of the cell by dynamic secondary ion mass spectrometry (SIMS microscopy). Biochim. Biophys. Acta 1724, 228–238.

Hahn, A.T., Jones, J.T., and Meyer, T. (2009). Quantitative analysis of cell cycle phase durations and PC12 differentiation using fluorescent biosensors. Cell Cycle 8, 1044–1052.

Haubner, B.J., Adamowicz-Brice, M., Khadayate, S., Tiefenthaler, V., Metzler, B., Aitman, T., and Penninger, J.M. (2012). Complete cardiac regeneration in a mouse model of myocardial infarction. Aging (Albany, N.Y.) 4, 966–977.

Herget, G.W., Neuburger, M., Plagwitz, R., and Adler, C.P. (1997). DNA content, ploidy level and number of nuclei in the human heart after myocardial infarction. Cardiovasc. Res. 36, 45–51.

Jesty, S.A., Steffey, M.A., Lee, F.K., Breitbach, M., Hesse, M., Reining, S., Lee, J.C., Doran, R.M., Nikitin, A.Y., Fleischmann, B.K., and Kotlikoff, M.I. (2012). c-kit+ precursors support postinfarction myogenesis in the neonatal, but not adult, heart. Proc. Natl. Acad. Sci. USA 109, 13380–13385.

Lechene, C., Hillion, F., McMahon, G., Benson, D., Kleinfeld, A.M., Kampf, J.P., Distel, D., Luyten, Y., Bonventre, J., Hentschel, D., et al. (2006). High-resolution quantitative imaging of mammalian and bacterial cells using stable isotope mass spectrometry. J. Biol. 5, 20.

Leu, M., Ehler, E., and Perriard, J.C. (2001). Characterisation of postnatal growth of the murine heart. Anat. Embryol. (Berl.) 204, 217–224.

Mahmoud, A.I., Kocabas, F., Muralidhar, S.A., Kimura, W., Koura, A.S., Thet, S., Porrello, E.R., and Sadek, H.A. (2013). Meis1 regulates postnatal cardiomyocyte cell cycle arrest. Nature 497, 249–253.

Mayhew, T.M., Pharaoh, A., Austin, A., and Fagan, D.G. (1997). Stereological estimates of nuclear number in human ventricular cardiomyocytes before and after birth obtained using physical disectors. J. Anat. 191, 107–115.

Messaoudii, C., Boudier, T., Sanchez Sorzano, C.O., and Marco, S. (2007). TomoJ: tomography software for three-dimensional reconstruction in transmission electron microscopy. BMC Bioinformatics 8, 288.

Mollova, M., Bersell, K., Walsh, S., Savla, J., Das, L.T., Park, S.Y., Silberstein, L.E., Dos Remedios, C.G., Graham, D., Colan, S., and Kühn, B. (2013). Cardiomyocyte proliferation contributes to heart growth in young humans. Proc. Natl. Acad. Sci. USA 110, 1446–1451.

Naqvi, N., Li, M., Calvert, J.W., Tejada, T., Lambert, J.P., Wu, J., Kesteven, S.H., Holman, S.R., Matsuda, T., Lovelock, J.D., et al. (2014). A proliferative burst during preadolescence establishes the final cardiomyocyte number. Cell 157, 795–807.

Naqvi, N., Singh, R., Iismaa, S.E., Li, M., Calvert, J.W., Martin, D.I.K., Harvey, R.P., Graham, R.M., and Husain, A. (2015). Cardiomyocytes Replicate and their Numbers Increase in Young Hearts. Cell 163, this issue, 783–784.

Palpant, N.J., and Murry, C.E. (2014). Proliferation at the heart of preadolescence. Cell 157, 765–767.

Ponti, G., Obernier, K., Guinto, C., Jose, L., Bonfanti, L., and Alvarez-Buylla, A. (2013). Cell cycle and lineage progression of neural progenitors in the ventric-ular-subventricular zones of adult mice. Proc. Natl. Acad. Sci. USA 110, E1045–E1054.

Porrello, E.R., Mahmoud, A.I., Simpson, E., Hill, J.A., Richardson, J.A., Olson, E.N., and Sadek, H.A. (2011). Transient regenerative potential of the neonatal mouse heart. Science 331, 1078–1080.

Preissl, S., Schwaderer, M., Raulf, A., Hesse, M., Grüning, B.A., Köbele, C., Backofen, R., Fleischmann, B.K., Hein, L., and Gilsbach, R. (2015). Decipher-ing the Epigenetic Code of Cardiac Myocyte Transcription. Circ. Res. 117, 413–423. http://dx.doi.org/10.1161/CIRCRESAHA.115.306337.

Puente, B.N., Kimura, W., Muralidhar, S.A., Moon, J., Amatruda, J.F., Phelps, K.L., Grinsfelder, D., Rothermel, B.A., Chen, R., Garcia, J.A., et al. (2014). The oxygen-rich postnatal environment induces cardiomyocyte cell-cycle arrest through DNA damage response. Cell 157, 565–579.

Sadek, H.A., Martin, J.F., Takeuchi, J.K., Leor, J., Nie, Y., Giacca, M., and Lee, R.T. (2014). Multi-investigator letter on reproducibility of neonatal heart regen-eration following apical resection. Stem Cell Reports 3, 1.

Scholzen, T., and Gerdes, J. (2000). The Ki-67 protein: from the known and the unknown. J. Cell. Physiol. 182, 311–322.

Senyo, S.E., Steinhauser, M.L., Pizzimenti, C.L., Yang, V.K., Cai, L., Wang, M., Wu, T.D., Guerquin-Kern, J.L., Lechene, C.P., and Lee, R.T. (2013). Mamma-lian heart renewal by pre-existing cardiomyocytes. Nature 493, 433–436.

Soonpaa, M.H., and Field, L.J. (1998). Survey of studies examining mammalian cardiomyocyte DNA synthesis. Circ. Res. 83, 15–26.

Soonpaa, M.H., Kim, K.K., Pajak, L., Franklin, M., and Field, L.J. (1996). Cardi-omyocyte DNA synthesis and binucleation during murine development. Am. J. Physiol. 271, H2183–H2189.

Soonpaa, M.H., Zebrowski, D.C., Platt, C., Rosenzweig, A., Engel, F.B., and Field, L.J. (2015). Cardiomyocyte Cell-Cycle Activity during Preadolescence. Cell 163, this issue, 781–782.

Srsen, V., Fant, X., Heald, R., Rabouille, C., and Merdes, A. (2009). Centro-some proteins form an insoluble perinuclear matrix during muscle cell differen-tiation. BMC Cell Biol. 10, 28.

Steinhauser, M.L., Bailey, A.P., Senyo, S.E., Guillermier, C., Perlstein, T.S., Gould, A.P., Lee, R.T., and Lechene, C.P. (2012). Multi-isotope imaging mass spectrometry quantifies stem cell division and metabolism. Nature 481, 516–519.

Walsh, S., Pontén, A., Fleischmann, B.K., and Jovinge, S. (2010). Cardiomyo-cyte cell cycle control and growth estimation in vivo–an analysis based on car-diomyocyte nuclei. Cardiovasc. Res. 86, 365–373.

Wilson, A., Laurenti, E., Oser, G., van der Wath, R.C., Blanco-Bose, W., Jawor-ski, M., Offner, S., Dunant, C.F., Eshkind, L., Bockamp, E., et al. (2008). He-matopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell 135, 1118–1129.

Yeung, M.S., Zdunek, S., Bergmann, O., Bernard, S., Salehpour, M., Alkass, K., Perl, S., Tisdale, J., Possnert, G., Brundin, L., et al. (2014). Dynamics of oligodendrocyte generation and myelination in the human brain. Cell 159, 766–774.

Zhang, C.-H., and Kühn, B. (2014). Muscling up the heart: a preadolescent car-diomyocyte proliferation contributes to heart growth. Circ. Res. 115, 690–692.

Erratum

Regulators of Gut Motility Revealed by a Gnotobiotic Model of Diet-Microbiome Interactions Related to Travel

Neelendu Dey, Vitas E. Wagner, Laura V. Blanton, Jiye Cheng, Luigi Fontana, Rashidul Haque, Tahmeed Ahmed, and Jeffrey I. Gordon*
*Correspondence: jgordon@wustl.edu
http://dx.doi.org/10.1016/j.cell.2015.10.052

(Cell *163*, 95–107; September 24, 2015)

In the above article, we show that changes in diet composition affect gut motility in a microbiota-dependent manner. While describing the rationale for re-deriving $Ret^{+/-}$ mice as germ free on page 103, right column, lines 6–9, we erroneously indicated that conventionally raised wild-type ($Ret^{+/+}$) mice have slower transit times than their heterozygous ($Ret^{+/-}$) littermates. The correct sentence should have stated, "We found that conventionally raised wild-type ($Ret^{+/+}$) mice have significantly lower transit times (faster motility) than their conventionally raised heterozygous ($Ret^{+/-}$) littermates (p = 0.05, one-tailed Student's t test; as shown in Table S2F)." The published results of this experiment were correct, and this textual error does not affect the conclusions of the paper. The online version of the paper has been corrected to reflect this change. We apologize for any confusion that it may have caused.

SnapShot: T Cell Exhaustion

Kristen E. Pauken and E. John Wherry
Department of Microbiology and Institute for Immunology, University of Pennsylvania, Philadelphia, PA 19104, USA

Property	Naive	Effector		Memory				Exhausted	
	Quiescent cells circulating through lymphoid tissues and blood	Generated following priming in the presence of antigen, costimulation, and inflammation		Common following acute infection and vaccination				Common during chronic infections and cancer due to persistent stimulation	
Subsets		KLRG1+ CD127-	KLRG1- CD127+	T_{SCM}	T_{CM}	T_{EM}	T_{RM}	Progenitor	Terminal Progeny
Markers	CD44low CD62L+	CD44high, CD62L-, Granzyme B, Perforin		CD95 Sca1	CD62L+	CD62L-	CD69+ CD103+	T-bethi PD-1int	Eomeshi PD-1hi
Proliferative capacity	+++	+/-		++++	+++	+	-	+/-	-
Effector capacity	-	+++		+/-	++	+++	+++	+/-	+/-
Location	LN, spleen, blood	Spleen, blood, peripheral tissues		LN, spleen, blood	LN, spleen, blood	Spleen, blood, non-lymphoid tissues	Non-lymphoid tissues	Spleen, blood	Spleen, blood, non-lymphoid tissues

ORIGINS OF EXHAUSTION

Naive → Effector
KLRG1+ CD127- CD127+ KLRG1-

Ag cleared — Inflammation resolves → **Memory** (T_{SCM}, T_{CM}, T_{EM}, T_{RM})

CD127+ KLRG1-

Ag persists — Chronic inflammation → **Exhausted** (Progenitor T_{EX}, Terminal T_{EX})

SUBSETS OF T_{EX}

Antigen-driven conversion through proliferation

T-bet → Progenitor subset → Eomes → Terminal subset

	Progenitor subset Tbethi PD-1int	Terminal subset Eomeshi PD-1hi
Defining markers		
Proliferative reserve	+	-
IFNγ/TNFα	+	+/-
Cytotoxicity	-	+/-
Localization	Spleen, blood	Spleen, blood, NLT
Reinvigoration with PD1 pathway blockade	++	-

SIGNALS FOR THE DEVELOPMENT OF T_{EX}

Drivers of Exhaustion

Soluble factors
Inflammatory cytokines (IFN-I, IL-6, IL-27)
Suppressive cytokines (IL-10, TGF-β)

Tissue microenvironment
Hypoxia, nutrients, pH

Persistent antigen

Over-expression of multiple inhibitory receptors
PD-1, Tim-3, Lag-3, TGIT, etc.

Regulatory leukocytes
Tregs, APCs

Physiological consequences

CD8 T cell — TIGIT, Tim-3, TCR, PD-1, Lag-3
Altered metabolism
Altered usage of transcription factors
Altered transcriptional program

Altered homeostasis: Poor ag-independent renewal, weak IL-7/IL-15 responsiveness

Cell death due to overstimulation

Poor control of infection or tumor

Progressive loss of effector functions (IFNγ, TNFα, IL-2, cytotoxicity)

HETEROGENEITY IN T_{EX} BASED ON LOCATION

Non-lymphoid sites of infection — Intestine — Draining lymph nodes — Liver — Spleen

Chronic infection — Cancer

Spleen — Tumor — Draining lymph nodes

Legend:
T-bethi progenitor cells
Heterogenous Eomeshi terminal cells
Tissue- or tumor-associated dendritic cell
Tissue- or tumor-associated macrophage
Blood vessel
Lymphatics

See online version for legend and references.

SnapShot: T Cell Exhaustion

Cell

Kristen E. Pauken and E. John Wherry
Department of Microbiology and Institute for Immunology, University of Pennsylvania, Philadelphia, PA 19104, USA

T_{EX} Are a Discrete T Cell Subset

T cell exhaustion is a distinct differentiation state that can be distinguished from naive, effector, and memory T cells. Compared to effector (T_E) and memory (T_{MEM}) T cells, exhausted T cells (T_{EX}) display impaired effector functions (e.g., rapid production of effector cytokines, cytotoxicity) (Wherry and Kurachi, 2015). T_{EX} have limited proliferative potential, especially compared to some subsets of T_{MEM} (e.g., T_{SCM} and T_{CM}) and naive T cells (T_N). T_{EX} also develop a distinct transcriptional program that is unlike T_N, T_E, or T_{MEM} (Doering et al., 2012; Martinez et al., 2015; Paley et al., 2012), a feature shared between chronic infections and cancer (Baitsch et al., 2011). T_{EX} confer only weak or temporary immune pressure on chronic infections or tumors, and this immunity is ultimately ineffective (Wherry and Kurachi, 2015). T_{EX} can be re-invigorated by blockade of PD-1 and other inhibitory receptors and immunoregulatory pathways (Barber et al., 2006; Wherry and Kurachi, 2015). This re-invigoration demonstrates that at least a subset of T_{EX} retains residual protective potential and has been translated into impressive clinical results in cancer (Page et al., 2014). Exhaustion can develop in both CD4+ and CD8+ T cells, though CD8+ T cell exhaustion is better understood and is the focus of this Snapshot.

Origins of Exhaustion

Unlike other forms of T cell dysfunction (e.g., anergy), exhaustion is not induced at priming (Wherry and Kurachi, 2015). Longitudinal analyses support the notion of gradual induction of dysfunction. For example, the transcriptional profiles of virus-specific CD8+ T cells during an acutely resolving versus developing chronic viral infection are largely similar at days 6 and 8 post-infection (p.i.). At day 15 p.i., these transcriptional programs diverge toward memory or exhaustion (Doering et al., 2012; Wherry and Kurachi, 2015). Moreover, CD8+ T cells primed during a chronic infection can develop into functional T_{MEM} if removed from the chronic infection at day 8 p.i., but not day 15 or 30 p.i. (Angelosanto et al., 2012; Brooks et al., 2006). T_{EX} and T_{MEM} also both develop from the CD127+KLRG1− "memory precursor" subset of T_E, indicating a common developmental origin for T_{EX} and T_{MEM} (Angelosanto et al., 2012).

Signals for the Development of T_{EX}

While the development of T_{EX} remains incompletely understood, persisting and likely continuous rather than intermittent antigen stimulation appears to be a key signal driving exhaustion. However, other types of signals are also likely important. These include: proinflammatory (e.g., IFN-α/β, IL-6, IL-27) and suppressive (e.g., IL-10, TGF-β) cytokines, other regulatory leukocytes (e.g., regulatory T cells, immunoregulatory antigen presenting cells), and the tissue microenvironment (e.g., altered hypoxia, nutrients, pH) (Wherry and Kurachi, 2015). Together with chronic TCR engagement, these signals drive elevated and sustained expression of multiple inhibitory receptors (e.g., PD-1, Lag-3, Tim-3, etc.), altered use of key transcription factors (e.g., T-bet, Eomes, Blimp-1, NFAT/AP-1), changes in metabolism, and a transcriptional program distinct from other T cell differentiation states. Ultimately, these signals lead to progressive loss of effector functions, altered homeostasis compared to T_{MEM}, and cell death due to overstimulation. As a result, T cell exhaustion results in poor control of pathogens or tumors. However, the ability to partially reverse exhaustion through strategies including PD-1 pathway blockade suggests that T_{EX}, or at least a subset of this pool, are not terminally differentiated and can contribute to protective immunity if re-invigorated.

Subsets of T_{EX}

At least two subsets of T_{EX} exist based on expression of PD-1 and the T-box transcription factors T-bet and Eomesodermin (Eomes) (Blackburn et al., 2008; Paley et al., 2012). One subset expresses high T-bet and intermediate PD-1 (T-bet^hi PD-1^int), whereas the other subset expresses high Eomes and high PD-1 but lower T-bet (Eomes^hi PD-1^hi). These two subsets differ in key functions. The T-bet^hi PD-1^int subset retains moderate proliferative capacity and some potential to produce effector cytokines (e.g., IFNγ, TNFα) but has limited cytotoxicity. The Eomes^hi PD-1^hi subset produces less cytokines and has poor proliferative potential but retains partial cytotoxicity compared to the T-bet^hi PD-1^int subset (Paley et al., 2012). Only the T-bet^hi PD-1^int subset can be reinvigorated by PD-1:PD-L1 pathway blockade, an observation with implications for immunotherapy (Blackburn et al., 2008). A lineage relationship exists between these two subsets: the T-bet^hi PD-1^int subset divides in response to persisting antigen, giving rise to Eomes^hi PD-1^hi progeny. The Eomes^hi PD-1^hi cells are terminally differentiated and do not convert back to T-bet^hi PD-1^int cells (Paley et al., 2012). Thus, the T-bet^hi PD-1^int cells are referred to as the "progenitor subset" and the Eomes^hi PD-1^hi as the "terminal subset."

T_{EX} Based on Anatomical Location

There is likely additional heterogeneity in the pool of T_{EX} based on tissue location and disease context. In chronic infection, the T-bet^hi progenitor subset is found in the spleen and blood, while the Eomes^hi PD-1^hi terminal subset is more abundant in non-lymphoid tissues (Paley et al., 2012) and possibly tumors. However, the distinct microenvironments of non-lymphoid tissues and tumors are likely to influence T_{EX} biology. Indeed, different tissues have different ratios of T-bet^hi to Eomes^hi subsets, and T_{EX} in these locations may differ in other key phenotypic and functional attributes (Blackburn et al., 2008; Paley et al., 2012). Therefore, while T_{EX} represents a discrete differentiation state compared to T_N, T_E, and T_{MEM}, the disease context and location may impart additional layers of heterogeneity.

REFERENCES

Angelosanto, J.M., Blackburn, S.D., Crawford, A., and Wherry, E.J. (2012). J. Virol. *86*, 8161–8170.

Baitsch, L., Baumgaertner, P., Devêvre, E., Raghav, S.K., Legat, A., Barba, L., Wieckowski, S., Bouzourene, H., Deplancke, B., Romero, P., et al. (2011). J. Clin. Invest. *121*, 2350–2360.

Barber, D.L., Wherry, E.J., Masopust, D., Zhu, B., Allison, J.P., Sharpe, A.H., Freeman, G.J., and Ahmed, R. (2006). Nature *439*, 682–687.

Blackburn, S.D., Shin, H., Freeman, G.J., and Wherry, E.J. (2008). Proc. Natl. Acad. Sci. USA *105*, 15016–15021.

Brooks, D.G., McGavern, D.B., and Oldstone, M.B. (2006). J. Clin. Invest. *116*, 1675–1685.

Doering, T.A., Crawford, A., Angelosanto, J.M., Paley, M.A., Ziegler, C.G., and Wherry, E.J. (2012). Immunity *37*, 1130–1144.

Martinez, G.J., Pereira, R.M., Äijö, T., Kim, E.Y., Marangoni, F., Pipkin, M.E., Togher, S., Heissmeyer, V., Zhang, Y.C., Crotty, S., et al. (2015). Immunity *42*, 265–278.

Page, D.B., Postow, M.A., Callahan, M.K., Allison, J.P., and Wolchok, J.D. (2014). Annu. Rev. Med. *65*, 185–202.

Paley, M.A., Kroy, D.C., Odorizzi, P.M., Johnnidis, J.B., Dolfi, D.V., Barnett, B.E., Bikoff, E.K., Robertson, E.J., Lauer, G.M., Reiner, S.L., and Wherry, E.J. (2012). Science *338*, 1220–1225.

Wherry, E.J., and Kurachi, M. (2015). Nat. Rev. Immunol. *15*, 486–499.